GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 31, 2003, 16:22:15 ; Search time 2303 Seconds
(without alignments)
12257.374 Million cell updates/sec

Title: US-10-069-541-5
Perfect score: 1743
Sequence: 1 atggctttccatgtggaagg.....ctgaagataatttacagtga 1743

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	518.8	29.8	672	17	AG157499	AG157499 Pan trogl
2	404	23.2	516	10	BE233479	BE233479 139685 MA
3	329.8	18.9	650	10	BB626260	BB626260 BB626260
4	312.8	17.9	541	10	AW668962	AW668962 111664 MA
5	274.2	15.7	524	12	BE723927	BE723927 198406 MA
6	263.2	15.1	800	9	AL669749	AL669749 AL669749
7	212.4	12.2	1037	9	AL666817	AL666817 AL666817
8	209.2	12.0	641	13	BI630566	BI630566 RH59836.5
9	207.4	11.9	640	13	BI629504	BI629504 RH58381.5
10	205.2	11.8	658	13	BM629925	BM629925 170006875
11	203.2	11.7	652	10	BB626456	BB626456 BB626456
12	186.4	10.7	624	13	BJ122485	BJ122485 BJ122485
13	179.8	10.3	580	13	BJ125564	BJ125564 BJ125564
14	166	9.5	604	9	AU199794	AU199794 AU199794
15	163.4	9.4	500	10	AV994375	AV994375 AV994375
16	163.2	9.4	555	13	BJ117801	BJ117801 BJ117801
c 17	155	8.9	561	17	AQ316435	AQ316435 RPCI11-10
18	153	8.8	584	13	BJ105382	BJ105382 BJ105382
c 19	134	7.7	632	17	AZ612750	AZ612750 1M0439J17
20	128.2	7.4	500	13	BJ105730	BJ105730 BJ105730
c 21	127.6	7.3	618	17	AZ908709	AZ908709 RPCI-24-2
22	121	6.9	926	17	CNS04L3J	AL295624 Tetraodon
c 23	117.2	6.7	525	13	BI508286	BI508286 BB170004A
c 24	117.2	6.7	530	13	BI503332	BI503332 BB170012A
25	112.8	6.5	355	9	AU209671	AU209671 AU209671
c 26	112.4	6.4	558	13	BI507950	BI507950 BB170010A
c 27	108	6.2	420	13	BI506529	BI506529 BB170027B
c 28	106.8	6.1	539	13	BI508323	BI508323 BB170007A
29	99.4	5.7	486	12	BG353180	BG353180 ps23d03.y
c 30	94.6	5.4	563	13	BI503188	BI503188 BB170001A
31	91.6	5.3	1058	17	CNS05R6F	AL350160 Tetraodon
32	91.4	5.2	1101	17	CNS05RBW	AL350357 Tetraodon
33	90.4	5.2	627	10	BB663292	BB663292 BB663292
c 34	84.4	4.8	936	17	CNS03C0M	AL237199 Tetraodon
c 35	81.4	4.7	741	13	BJ153899	BJ153899 BJ153899
36	77.8	4.5	367	9	AL839758	AL839758 AL839758
c 37	77.6	4.5	292	13	BI502965	BI502965 BB170004A
c 38	76.6	4.4	760	17	CNS033GD	AL226102 Tetraodon
39	73.2	4.2	962	17	CNS033GC	AL226101 Tetraodon
40	73	4.2	311	9	AL839681	AL839681 AL839681
c 41	68.2	3.9	696	17	CNS03MCS	AL250597 Tetraodon
42	67.4	3.9	807	17	CNS024XK	AL181361 Tetraodon
43	66	3.8	360	14	C13840	C13840 C13840 Yuji
c 44	64.6	3.7	1083	17	CNS051WV	AL317416 Tetraodon
45	64.2	3.7	984	17	CNS03C0L	AL237198 Tetraodon

ALIGNMENTS

RESULT 1
AG157499
LOCUS AG157499 672 bp DNA linear GSS 09-JAN-2002
DEFINITION Pan troglodytes DNA, clone: RP43-022H02.T7, genomic survey sequence.
ACCESSION AG157499
VERSION AG157499.1 GI:16687177
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee Male BAC Library clone:RP43-022H02.T7.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library RPCI-43
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 672)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library RPCI-43 This BAC end was generated during the R&D process and may have higher chance

```
          of clone tracking errors.
           PRIMERS
               Sequencing: T7
           LIBRARY
               Vector    : pBACe3.6
               R.Site 1  : EcoRI
               R.Site 2  : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .672
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="RP43-022H02.T7"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT       186 a    144 c    135 g    206 t      1 others
ORIGIN

  Query Match             29.8%;  Score 518.8;  DB 17;  Length 672;
  Best Local Similarity   97.3%;  Pred. No. 1.9e-134;
  Matches  549;  Conservative    0;  Mismatches   12;  Indels    3;  Gaps    2;

Qy   1110 AAATGCTTCGGACAAAGAAATCGTTTGGGTTATGCGAATCACAGTGTTTGTGTTTGGAGC 1169
           | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    111 ACAGGCTTCGGACAAAGAAATCGTTTGGGTTATGCGAATCACAGTGTTTGTGTTTGGAGC 170

Qy   1170 ATCTGCAACAGCCATGGCCTTGCTGACGAAAACTGTGTATGGGCTCTGGTACCTCAGTTC 1229
           |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db    171 ATCTGCAACAGCCATGGCCTTGCTGACGAAGACTGTGTATGGGCTCTGGTACCTCAGTTC 230

Qy   1230 TGACCTTGTTTACATCGTTATCTTCCCCCAGCTGCTTTGTGTACTCTTTGTTAAGGGAAC 1289
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    231 TGACCTTGTTTACATCGTTATCTTCCCCCAGCTGCTTTGTGTACTCTTTGTTAAGGGAAC 290

Qy   1290 CAACACCTATGGGGCCGTGGCAGGTTATGTTTCTGGCCTCTTCCTGAGAATAACTGGAGG 1349
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291 CAACACCTATGGGGCCGTGGCAGGTTATGTTTCTGGCCTCTTCCTGAGAATAACTGGAGG 350

Qy   1350 GGAGCCATATCTGTATCTTCAGCCCTTGATCTTCTACCCTGGCTATTACCCTGATGATAA 1409
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    351 GGAGCCATATCTGTATCTTCAGCCCTTGATCTTCTACCCTGGCTATTACCCTGATGATAA 410

Qy   1410 TGGTATATATAATCAGAAATTTCCATTTAAAACACTTGCCATGGTTACATCATTCTTAAC 1469
           |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    411 TGGTATATATAATCAGAAATTTCCATTTAAAACACTTGCCATGGTTACGTCATTCTTAAC 470

Qy   1470 CAACATTTGCATCTCCTATCTAGCCAAGTATCTATTTG-AAAGTGGAACCTTGCCACCTA 1528
           |||||||||| |||||||||||||||| |||||||||| ||||||||||||||||||||
Db    471 CAACATTTGCGTCTCCTATCTAGCCAAATATCTATTTGAAAAGTGGAACCTTGCCACCTA 530

Qy   1529 AATTAGATGTATTTGATGCTGTTGTTGCAAGACACAGTGAAGAAAACATGGATAAGACAA 1588
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    531 AATTAGATGTATTTGATGCTGTTGTTGCAAGACACAGTGAAGAAAACATGGATAAGACAA 590

Qy   1589 TTCTTGTCAAAAATGAAAATATTAAATTAGATGAACTTGCACTTGTGAAGCCACGACAGA 1648
           || |||||||||||||||  |||||||||||| ||||||||||||||||||||||||||
Db    591 TTTTTGTCAAAAATGAAA--TATAAATTAGATGACCTTGCACTTGTGAAGCCACGACAGA 648

Qy   1649 GCATGACCCTCAGCTCAACTTTCA 1672
           ||||||||||||||| ||||||||
Db    649 ACATGACCCTCAGCTTAACTTTCA 672


RESULT 2
BE233479
LOCUS       BE233479                 516 bp    mRNA    linear   EST 10-JUL-2000
DEFINITION  139685 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BE233479
VERSION     BE233479.1  GI:9018197
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 516)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 75  row: G  column: 12
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .516
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 1PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from day 11, 13, 15, 20,
                     and 30 embryos."
BASE COUNT       128 a    129 c    112 g    147 t
ORIGIN

  Query Match             23.2%;  Score 404;  DB 10;  Length 516;
  Best Local Similarity   86.4%;  Pred. No. 3.1e-102;
  Matches  446;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy   1067 GTTCCATGTTTGCACGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAG 1126
           |||| ||||||||  | |||||||| ||||| || |||||||||||| ||||| ||||  |
Db      1 GTTCTATGTTTGCTAGAAACATCTAGCAGCTCTCATTCAGACAAAACGCTTCCGACAGGG 60

Qy   1127 AAATCGTTTGGGTTATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGG 1186
           | ||||| ||||| ||||| |||||||| || |||||||| || ||||||||||||||||
Db     61 AGATCGTCTGGGTCATGCGGATCACAGTATTCGTGTTTGGTGCGTCTGCAACAGCCATGG 120

Qy   1187 CCTTGCTGACGAAAACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCG 1246
           |||||||||| || ||||||||||||||||||||||||||| || ||||| || ||||||
Db    121 CCTTGCTGACCAAGACCGTGTATGGGCTCTGGTACCTCAGCTCCGACCTCGTCTACATCA 180

Qy   1247 TTATCTTCCCCCAGCTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCG 1306
           |||||||||| |||||||| ||||| |||||  | ||||| |||||||| || |||||||
Db    181 TTATCTTCCCGCAGCTGCTCTGTGTGCTCTTCATCAAGGGGACCAACACGTACGGGGCCG 240

Qy   1307 TGGCAGGTTATGTTTCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATC 1366
           ||||||| ||  || |||||||| ||||||||  |||| || || |||||||| ||| | |
Db    241 TGGCAGGGTACATTGCTGGCCTTTTCCTGAGGGTAACCGGTGGAGAGCCATACCTGAACC 300

Qy   1367 TTCAGCCCTTGATCTTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGA 1426
           | |||||||||||||| |||||||| ||||||  |||  | ||||||||||||||||||
Db    301 TGCAGCCCTTGATCTTTTACCCTGGTTATTACGTTGAAAAAAATGGTATATATAATCAGA 360

Qy   1427 AATTTCCATTTAAAACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCT 1486
           ||| |||||||||| ||||||||| | || || |||||| |||||||||||||||| |
Db    361 GATTCCCATTTAAAACCCTTGCCATGCTCACCTCCTTCTTATCCAACATTTGCATCTCTT 420

Qy   1487 ATCTAGCCAAGTATCTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATG 1546
           |||||||||| |||||||||||||||||||||||||||||||| ||||||||| | |||||||
Db    421 ATCTAGCCAAATATCTATTTGAAAGTGGAACCTTGCCACCAAAATTAGATATGTTTGATG 480

Qy   1547 CTGTTGTTGCAAGACACAGTGAAGAAAACATGGATA 1582
           |||||||||||||||||||||||||||||||||||
```

```
Db    481 CTGTTGTTGCAAGACACAGTGAAGAAAACATGGATA 516


RESULT 3
BB626260
LOCUS       BB626260                650 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB626260 RIKEN full-length enriched, adult male diencephalon Mus
            musculus cDNA clone 9330170D24 5', mRNA sequence.
ACCESSION   BB626260
VERSION     BB626260.1  GI:16464298
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 650)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .650
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="9330170D24"
                     /clone_lib="RIKEN full-length enriched, adult male
                     diencephalon"
                     /sex="male"
                     /tissue_type="diencephalon"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 185.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
BASE COUNT      139 a    168 c    174 g    165 t      4 others
ORIGIN

  Query Match           18.9%;  Score 329.8;  DB 10;  Length 650;
  Best Local Similarity  86.0%;  Pred. No. 2.3e-81;
  Matches  375;  Conservative    0;  Mismatches   60;  Indels    1;  Gaps    1;

Qy      1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60
          ||| ||||||| || |||||||||| ||||||| ||| | |||||||| || || || |||
Db    201 ATGTCTTTCCACGTAGAAGGACTGGTAGCTATTATCCTCTTCTACCTCCTTATATTTCTG 260

Qy     61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120
          |||||||||||||||| |||| |||||||||||| |||| | ||||||||||||| |||
Db    261 GTTGGAATATGGGCTGCATGGAAAACCAAAAACAGCGGCAACCCAGAAGAGCGCAGTGAA 320

Qy    121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180
          |||||||||| || ||||| || |||||||| |||||||||| ||||||||||||||
Db    321 GCCATCATAGTCGGGGGCCGTGACATTGGTTTGTTGGTTGGTGGTTTTACCATGACAGCC 380

Qy    181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240
          |||||||| |||||||| || |||||||| ||||| |||||||| |||| |||||||| |
Db    381 ACCTGGGTTGGAGGAGGCTACATCAATGGGACAGCAGAAGCAGTGTATGGGCCAGGTTGT 440

Qy    241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300
          || |||||||| ||||| ||||| |||||||||||||| ||||| |||||||||| |||
Db    441 GGTCTAGCTTGNGCTCANGCACCCATTGGATATTCTCTGAGTCTAATTTTAGGTGGTCTG 500

Qy    301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360
          || |||||  |||||||||  || ||||| |||||||| |||||||||| ||||| ||
Db    501 TTTTTTGC-GAACCTATGCNGTCCAAGGGATATGTGACTATGTTAGACCCATTTCAACAG 559

Qy    361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420
          |||||||||| |||||||| || || || || || ||||| |||||||||| ||||||
Db    560 ATCTATGGAAAGCGCATGGGTGGGCTGCTCTTCATCCCTGCCCTGATGGGAGAGATGTTC 619

Qy    421 TGGGCTGCAGCAATTT 436
          |||||||||| ||||
Db    620 TGGGCTGCAGCCATTT 635


RESULT 4
AW668962
LOCUS       AW668962                541 bp    mRNA    linear   EST 25-APR-2001
DEFINITION  111664 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW668962
VERSION     AW668962.1  GI:7525476
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 541)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
```

```
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 95  row: L  column: 20
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .541
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 1BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from lymph node, ovary,
                     fat, hypothalamus, and pituitary."
BASE COUNT      117 a    132 c    127 g    165 t
ORIGIN

  Query Match             17.9%;  Score 312.8;  DB 10;  Length 541;
  Best Local Similarity   85.0%;  Pred. No. 1.3e-76;
  Matches  350;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy     889 TCAACAGACTGGAACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCA 948
           ||||||| |||||||||||||||||||| |||| ||  | ||||| |   | | | ||||||
Db     130 TCAACAGCCTGGAACCAGACTGCATACGGGCCTCTTGCTCCCAGGGAGAAACAGGAGGCA 189

Qy     949 GACATGATTTTACCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGT 1008
           |||||||| || || ||||| ||  |||||||||||| ||||| ||||||| |||||||
Db     190 GACATGATCTTGCCGATTGTCCTCAAGTATCTCTGCCCCGTGTACATTTCTTACTTTGGT 249

Qy    1009 CTTGGTGCAGTTTCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGT 1068
           ||||| || |||||||||||||| ||||| ||||||||||||||||||||||||||||||
Db     250 CTTGGAGCCGTTTCTGCTGCTGTCATGTCCTCAGCAGATTCTTCCATCTTGTCAGCAAGT 309

Qy    1069 TCCATGTTTGCACGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAA 1128
           || |||||||| || |||||||||||||||||| |||||||||||||||| ||||| ||
Db     310 TCGATGTTTGCTCGCAACATCTACCAGCTTTCATTCAGACAAAATGCTTCTGACAAGGAG 369

Qy    1129 ATCGTTTGGGTTATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCC 1188
           || || ||||| ||||| ||||| || |||||||||||||| ||||| |   ||||||||
Db     370 ATAGTCTGGGTCATGCGCATCACGGTATTTGTGTTTGGAGCTTCTGCGATGACCATGGCC 429

Qy    1189 TTGCTGACGAAAACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTT 1248
           ||||| ||||| || |||||||||||||||||||||| |||||||| || |||||| |
Db     430 TTGCTAACGAAGACGGTGTATGGGCTCTGGTACCTCAGCTCTGACCTGGTCTACATCATC 489

Qy    1249 ATCTTCCCCCAGCTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATG 1300
           |||||||| ||| |||| || || |||||  | ||||| |||||||| ||||
Db     490 ATCTTCCCGCAGTTGCTCTGCGTGCTCTTCATCAAGGGTACCAACACGTATG 541


RESULT 5
BE723927
LOCUS       BE723927                 524 bp    mRNA    linear   EST 25-APR-2001
DEFINITION  198406 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE723927
VERSION     BE723927.1  GI:10125223
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 524)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 106  row: L  column: 14
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .524
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 4BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from day 20 and day 40
                     embryos."
BASE COUNT      111 a    129 c    118 g    166 t
ORIGIN

  Query Match             15.7%;  Score 274.2;  DB 12;  Length 524;
  Best Local Similarity   84.7%;  Pred. No. 9e-66;
  Matches  331;  Conservative    0;  Mismatches   58;  Indels    2;  Gaps    2;

Qy     889 TCAACAGACTGGAACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCA 948
           |||||| |||||||||||||||||| |||| ||  | ||||| |   | | | ||||||
Db     130 TCAACATCCTGGAACCAGACTGCATACGGGCCTCTTGCTCCCAGGGAGAAACAGGAGGCA 189

Qy     949 GACATGATTTTACCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGT 1008
           |||||||| || || ||||| ||  |||||||||||| ||||| ||||||| |||||||
Db     190 GACATGATCTTGCCGATTGTCCTCAAGTATCTCTGCCCCGTGTACATTTCTTACTTTGGT 249

Qy    1009 CTTGGTGCAGTTTCTGCTGCTGTTATGTCATCAGCAGATTC-TTCCATCTTGTCAGCAAG 1067
           ||||| || |||||||||||||| ||||| |||||||||| ||||||||||||||||||
Db     250 CTTGGAGCCGTTTCTGCTGCTGTCATGTCCTCAGCAGATTCTTTCCATCTTGTCAGCAAG 309

Qy    1068 TTCCATGTTTGCACGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGA 1127
           ||| |||||||| || |||||||||||||||| ||||||||||||||||| |||||| ||
Db     310 TTCGATGTTTGCTCGCAACATCTACCAGCTTTCATTCAGACAAAATGCTTCTGACAAGGA 369

Qy    1128 AATCGTTTGGGTTATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGC 1187
             || || ||||| ||||| ||||| || |||||||||||||| ||||| |   |||||||
Db     370 GATAGTCTGGGTCATGCGCATCACGGTATTTGTGTTTGGAGCTTCTGCGATGACCATGGC 429

Qy    1188 CTTGCTGACGAAAACTGTGTATGGGCTC-TGGTACCTCAGTTCTGACCTTGTTTACATCG 1246
           |||||| ||||| || |||||||||||| |||||||||| |||||||| || ||||||
Db     430 CTTGCTAACGAAGACGGTGTATGGGCTCTTGGTACCTCAGCTCTGACCTGGTCTACATCA 489

Qy    1247 TTATCTTCCCCCAGCTGCTTTGTGTACTCTT 1277
           | |||||||| ||| |||| || || |||||
Db     490 TCATCTTCCCGCAGTTGCTCTGCGTGCTCTT 520
```

```
RESULT 6
AL669749
LOCUS       AL669749                800 bp    mRNA    linear   EST 14-JAN-2002
DEFINITION  AL669749 directional larval cDNA library Ciona intestinalis cDNA
            clone 052ZB03 5', mRNA sequence.
ACCESSION   AL669749
VERSION     AL669749.1  GI:18143007
KEYWORDS    EST.
SOURCE      Ciona intestinalis.
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 800)
  AUTHORS   Genoscope.
  TITLE     Ciona intestinalis directional larval cDNA library
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            IMPORTANT: this sequence may contain errors. The Ciona intestinalis
            library from which the clone was isolated may be contaminated with
            cDNAs from bacteria or other Eukarya.
            Directional larval cDNA library originate from Dr.M.Branno,
            Stazione A.Dohrn, Naples, Italy, and was prepared in
            pBluescript2SK+.
FEATURES             Location/Qualifiers
     source          1. .800
                     /organism="Ciona intestinalis"
                     /db_xref="taxon:7719"
                     /clone="052ZB03"
                     /clone_lib="directional larval cDNA library"
                     /note="Vector: pBluescript2SK+"
BASE COUNT      171 a    181 c    183 g    257 t      8 others
ORIGIN

  Query Match             15.1%;  Score 263.2;  DB 9;  Length 800;
  Best Local Similarity   62.7%;  Pred. No. 1.2e-62;
  Matches  504;  Conservative    0;  Mismatches  286;  Indels   14;  Gaps    6;

Qy    250 TGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTGTTCTTTGC- 308
          ||| | || ||||| |||||||| ||       | ||  | || ||| | ||||||||
Db      2 TGGACGCAAGCACCCATTGGATACGCTTGCGCGTTAATACTTGGCGGCTTATTCTTTGCG 61

Qy    309 AAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAAATCTATGG 368
          ||     |||||      |||| |||||||| ||||| || ||  | ||||  | || |||
Db     62 AAGTAAAATGCGAAGTGAGGGATATGTGACGATGTTGGATCCACTGCAGCGCAACT-TGG 120

Qy    369 AAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTCTGGGCTGC 428
            |   ||||| |   | || | ||| ||||||||    |||||| | |||||| ||||
Db    121 TCGAGTAATGGGAGCGTTTCTTTATATACCTGCACTTGCTGGAGAATTATTCTGGTCTGC 180

Qy    429 AGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGATATGCACAT 488
          ||| || ||  | || ||||| |  ||| ||| || ||||| ||| |  ||||
Db    181 AGCTATATTGGCCGCGTTGGGCGGTACCTTCA-TGTTATCATTGATCTTCATATAACTGC 239

Qy    489 TTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGGCTCTATTC 548
            |||| ||  | ||||||  ||||||   | | |||||  |||  || || || || ||
Db    240 AGCTGTAATAGTATCTGCATGCATTGCTGTTGTATACACCATGGCCGGTGGTCTTTACTC 299

Qy    549 TGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGGATCAGCGT 608
          || || || || ||||| |||||| |  ||||||| ||  | || |||||| | ||| |
Db    300 GGTTGCTTATACAGATGTAGTTCAGTTGATTTGCATATTCATTGGACTGTGGTTGAGCAT 359

Qy    609 CCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTGCATGCCAA 668
          || || || ||  | |||||||| || |||||||||     || |||   ||  |
Db    360 TCCATTCGCGTTCACTCATCCTGCTGTATCAGACATCGCCACTACAGCTTACCACTCACC 419

Qy    669 ATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGGCTTGATAG 728
            ||         ||||| || |||   ||    ||                ||| | ||
Db    420 TAAC---------TGGCTTGGTACTTGGGATATTTCGACCACTGGTCTATGGATCGACTC 470

Qy    729 TTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGGGTTCTCTC 788
          |  |||| |  || |  | |||||||| || |||||||  |||||||| || ||| | ||
Db    471 TGCTCTGCTACTGTTATTTGGTGGAATACCGTGGCAAGTTTACTTTCAAAGAGTTTTATC 530

Qy    789 TTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGGTGCCTGGT 848
            |         |       |||||   ||| || ||| | || || ||||| ||  || |
Db    531 GNCNAAAAGNNCANGAAGCGCTCAGAAGCTTTCATTCATTGCTGCGTTCGGATGTTTGTT 590

Qy    849 GATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGGAACCAGAC 908
            ||| | || ||  | ||| | || || || |||| ||||| ||||| ||| ||   ||
Db    591 CATGTCAATACCTTCGATATTNATCGGTGCAATTGCTGCATCTACAGATTGGGACGCAAC 650

Qy    909 TGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTACCAATTGT 968
            | || || || |||  |||       | ||  ||   || ||  ||| |||| |||||
Db    651 ATCGTACGGCCTCCCAAGTCCAGTTGANAAAGGCGACCAAGCCAATATTCTACCCATTGT 710

Qy    969 TCTGCAGTATCTC-TGCCCTGTGTATATTTCTTTCTTT-GGTCTTGGTGCAGTTTCTGCT 1026
            || || || |||   ||||||   | | || |||||| || ||||| || |||||||||
Db    711 GCTTCAATACCTCACCCCCTGTAGCTGTATCATTCTTTGGGGCTTGGCGCTGTTTCTGCT 770

Qy   1027 GCTGTTATGTCATCAGCAGATTCT 1050
          ||||| |||||||| || || |||
Db    771 GCTGTNATGTCATCTGCCGACTCT 794


RESULT 7
AL666817
LOCUS       AL666817                1037 bp    mRNA    linear   EST 14-JAN-2002
DEFINITION  AL666817 directional larval cDNA library Ciona intestinalis cDNA
            clone 011ZE07 5', mRNA sequence.
ACCESSION   AL666817
VERSION     AL666817.1  GI:18133724
KEYWORDS    EST.
SOURCE      Ciona intestinalis.
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 1037)
  AUTHORS   Genoscope.
  TITLE     Ciona intestinalis directional larval cDNA library
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            IMPORTANT: this sequence may contain errors. The Ciona intestinalis
            library from which the clone was isolated may be contaminated with
            cDNAs from bacteria or other Eukarya.
            Directional larval cDNA library originate from Dr.M.Branno,
            Stazione A.Dohrn, Naples, Italy, and was prepared in
            pBluescript2SK+.
FEATURES             Location/Qualifiers
     source          1. .1037
                     /organism="Ciona intestinalis"
                     /db_xref="taxon:7719"
                     /clone="011ZE07"
                     /clone_lib="directional larval cDNA library"
                     /note="Vector: pBluescript2SK+"
BASE COUNT      280 a    199 c    216 g    339 t      3 others
ORIGIN

  Query Match             12.2%;  Score 212.4;  DB 9;  Length 1037;
  Best Local Similarity   57.8%;  Pred. No. 2.4e-48;
  Matches  434;  Conservative    0;  Mismatches  287;  Indels   30;  Gaps    2;

Qy    808 GCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGGTGCCTGGTGATGGCCATCCCAGCCATA 867
          |||||   ||| || ||| | || || ||||| ||  || | ||| | || ||  | |||
Db     18 GCTCAGAAGCTTTCATTCATTGCTGCGTTCGGATGTTTGTTCATGTCAATACCTTCGATA 77

Qy    868 CTCATTGGGGCCATTGGAGCATCAACAGACTGGAACCAGACTGCATATGGGCTTCCAGAT 927
            | || || || ||||  ||||| ||||| ||||||   ||  | || || || ||| |
Db     78 TTAATCGGTGCAATTGCTGCATCTACAGATTGGAACGCAACATCGTACGGCCTCCCAAGT 137
```

```
Qy     928 CCCAAGACTACAGAAGAGGCAGACATGATTTTACCAATTGTTCTGCAGTATCTCTGCCCT 987
           ||       | ||  ||    || ||  ||| |||| ||||| || || || |||  ||||
Db     138 CCAGTTGAAAAAGGCGACCAAGCCAATATTCTACCCATTGTGCTTCAATACCTCACCCCT 197

Qy     988 GTGTATATTTCTTTCTTTGGTCTTGGTGCAGTTTCTGCTGCTGTTATGTCATCAGCAGAT 1047
           ||   | | || |||||||| ||||| || ||||||||||||||| |||||||| || ||
Db     198 GTAGCTGTATCATTCTTTGGGCTTGGCGCTGTTTCTGCTGCTGTAATGTCATCTGCCGAC 257

Qy    1048 TCTTCCATCTTGTCAGCAAGTTCCATGTTTGCACGGAACATCTACCAGCTTTCCTTCAGA 1107
           |||   ||  | || ||    |   |||||| |  |||| || ||  |  |   | || ||
Db     258 TCTAGTATTCTATCTGCTTCTAGTATGTTTACTAGGAATATTTATNACGTCGTCATCCGA 317

Qy    1108 CAAAATGCTTCGGACAAAGAAATCGTTTGGGTTATGCGAATCACAGTGTTTGTGTTTGGA 1167
           | ||| || |||||     ||  | || ||||| ||| || |  || |  |  |  | |
Db     318 CCAAAGGCATCGGAGTTGGAGCTGGTGTGGGTAATGAGAGTATCAATAATATTCATAGCG 377

Qy    1168 GCATCTGCAACAGCCATGGCCTTGCTGACGAAAACTGTGTATGGGCTCTGGTACCTCAGT 1227
           |||   || || ||  | ||  |  |   |||| |||| |||   || |  || ||  ||
Db     378 GCAGGAGCCACTGCTTTAGCTCTTGTTGTGAAATCTGTTTATTATCTTTTCTATCTTTGT 437

Qy    1228 TCTGACCTTGTTTACATCGTTATCTTCCCCCAGCTGCTTTGTGTACTCTTTGTTAAGGGA 1287
           || |||||  | ||| || |  | || || ||| |  |  |||| ||||   |   ||
Db     438 TCGGACCTGATCTACGTCATACTTTTTCCACAGTTTGTAAGTGTTCTCTACCT---GGAC 494

Qy    1288 ACCAACACCTATGGGGCCGTGGCAGGTTATGTTTCTGGCCTCTTCCTGAGAATAACTGGA 1347
           | || || |||||  |  |    | || | |    ||  |  |  || |    ||
Db     495 CCAAATACGTATGGATCGATTCTTGCTTTTCTGGTGGGTTTGGTTTTACGACTTGGGGGT 554

Qy    1348 GGGGAGCCATATCTGTATCTTCAGCCCTTGATCTTCTACCCTGGCTATTACCCTGATGAT 1407
           || |||||      |      |||| || || |||   |||||
Db     555 GGCGAGCCTGGGTTTGGGTTTCAACCTTTTATCAAATACCC------------------- 595

Qy    1408 AATGGTATATATAATCAGAAATTTCCATTTAAAACACTTGCCATGGTTACATCATTCTTA 1467
                |||       || ||||||||||||||||||| |||||  |  ||  |  ||
Db     596 --------ATATGGAAGCAACTTTCCATTTAAAACACTTGCAATGGTATCTTCCATTGTA 647

Qy    1468 ACCAACATTTGCATCTCCTATCTAGCCAAGTATCTATTTGAAAGTGGAACCTTGCCACCT 1527
           ||      |      |||| ||  | |    || | ||||||   |||| |   | ||| |
Db     648 ACTCTGCTGGTTGTCTCTTACTTCGTTCGGTTTTTATTTGTCCGTGGTATAATACCAGCG 707

Qy    1528 AAATTAGATGTATTTGATGCTGTTGTTGCAA 1558
           | ||  ||| | ||  |      | ||||||
Db     708 AGATACGATTTCTTGAAATGCAATCTTGCAA 738
```

RESULT 8
BI630566

```
LOCUS       BI630566                 641 bp    mRNA    linear   EST 10-SEP-2001
DEFINITION  RH59836.5prime RH Drosophila melanogaster normalized Head pFlc-1
            Drosophila melanogaster cDNA clone RH59836 5 similar to CG7708:
            FBan0007708 GO:[] located on: 3R 91B8-91B8;: 08/23/2001, mRNA
            sequence.
ACCESSION   BI630566
VERSION     BI630566.1  GI:15532776
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 641)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
            ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
            ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
            Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
            ,G.M.
  TITLE     BDGP/HHMI RH Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003723: arm:3R [14365708,14621331]
            estimated-cyto:91B5-91D2: 08/23/2001
            Plate: RH.598  row: C  column: 12
            High quality sequence stop: 623.
FEATURES             Location/Qualifiers
     source          1. .641
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="RH59836"
                     /clone_lib="RH Drosophila melanogaster normalized Head
                     pFlc-1"
                     /sex="male and female"
                     /dev_stage="Adult"
                     /lab_host="DH5-alpha TonA"
                     /note="Organ: head; Vector: pFlc1; Site_1: XhoI; Site_2:
                     BamHI; Library was kindly generated by Piero Carninci at
                     the RIKEN. The library  was normalized and excised using
                     Cre recombinase. Plasmid cDNA library."
BASE COUNT      114 a    192 c    176 g    159 t
ORIGIN
```

Query Match 12.0%; Score 209.2; DB 13; Length 641;
Best Local Similarity 59.1%; Pred. No. 1.7e-47;
Matches 377; Conservative 0; Mismatches 258; Indels 3; Gaps 1;

```
Qy     769 TACTTTCAGAGGGTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTG 828
           ||||| ||| |||| || ||       | ||       || ||  | |||||||   |
Db       4 TACTTCCAGCGGGTGCTGTCCAGCAAAACGGCAGGAAGGGCCCAGCTTCTGTCCTATGTT 63

Qy     829 GCAGCTTTCGGGTGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCA 888
           |||||   ||| ||| |  |||||||||| ||  |  | |||||  || || ||||
Db      64 GCAGCCGCCGGATGCATTTTGATGGCCATTCCCCCGGTGCTCATCGGAGCGATTGCCAAG 123

Qy     889 TCAACAGACTGGAACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCA 948
           | |||   |||||| |||| |  ||   |  ||  |||||  |||    || ||| |
Db     124 GCTACACCTTGGAACGAGACAGATTACAAGGGACCCTATCCCCTGACCGTGGACGAGACG 183

Qy     949 GACATGATTTTACCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGT 1008
           |||||||  | || || || |||||||| |||   ||||  |   | || ||||||||
Db     184 AGCATGATTCTGCCCATGGTGCTGCAGTACCTCACGCCTGACTTCGTGTCCTTCTTTGGA 243

Qy    1009 CTTGGTGCAGTTTCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGT 1068
           | || || ||||| || || || ||||| || || || || ||  |  | || ||
Db     244 TTGGGCGCTGTTTCCGCCGCCGTGATGTCCTCCGCCGACTCCTCGGTGCTCTCCGCCGCC 303

Qy    1069 TCCATGTTTGCACGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAA 1128
           |||||||| || |||||| | ||| |  |   ||| | || || || || || || |  |||
Db     304 TCCATGTTCGCTCGGAACGTGTACAAATTGATTTTCCGTCAGAAGGCGTCCGAGATGGAA 363

Qy    1129 ATCGTTTGGGTTATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCC 1188
           |||  |||||||  |||||| || |  |  ||||| | ||        || ||   |||||||
Db     364 ATCATTTGGGTGATGCGAGTCGCCATCATTGTGGTGGGCATCCTGGCTACCATCATGGCC 423

Qy    1189 TTGCTGACGAAAACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTT 1248
           |   |      |  | || ||  | |||| | |  | || || || || ||| || ||
Db     424 CTCACCATTCCCTCCATCTACGGTTTGTGGTCCATGTGCTCGGATCTGGTCTACGTCATT 483

Qy    1249 ATCTTCCCCCAG---CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCC 1305
           | ||||| |||   ||| |    || | |||    |||     ||||| || ||    |
Db     484 CTGTTCCCGCAGCTACTGATGGTGGTGCACTTCAAGAAGCACTGCAACACGTACGGCAGC 543

Qy    1306 GTGGCAGGTTATGTTTCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTAT 1365
           || | |  ||  ||   | |||   | |  || |  | || || ||| |    ||
Db     544 CTGTCGGCATACATTGTGGCCCTGGCCATCCGACTGTCGGGCGGTGAGGCCATCTTGGGA 603

Qy    1366 CTTCAGCCCTTGATCTTCTACCCTGGCTATTACCCTGA 1403
           ||    || ||||||   || || |||||  ||   ||
Db     604 CTGGCTCCATTGATCAAGTATCCCGGCTACGACGAGGA 641
```

RESULT 9

```
BI629504
LOCUS       BI629504                640 bp    mRNA    linear   EST 10-SEP-2001
DEFINITION  RH58381.5prime RH Drosophila melanogaster normalized Head pFlc-1
            Drosophila melanogaster cDNA clone RH58381 5 similar to CG7708:
            FBan0007708 GO:[] located on: 3R 91B8-91B8;: 08/23/2001, mRNA
            sequence.
ACCESSION   BI629504
VERSION     BI629504.1  GI:15531714
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 640)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
            ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
            ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
            Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
            ,G.M.
  TITLE     BDGP/HHMI RH Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003723: arm:3R [14365708,14621331]
            estimated-cyto:91B5-91D2: 08/23/2001
            Plate: RH.583  row: G  column: 9
            High quality sequence stop: 512.
FEATURES             Location/Qualifiers
     source          1. .640
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="RH58381"
                     /clone_lib="RH Drosophila melanogaster normalized Head
                     pFlc-1"
                     /sex="male and female"
                     /dev_stage="Adult"
                     /lab_host="DH5-alpha TonA"
                     /note="Organ: head; Vector: pFlc1; Site_1: XhoI; Site_2:
                     BamHI; Library was kindly generated by Piero Carninci at
                     the RIKEN. The library  was normalized and excised using
                     Cre recombinase. Plasmid cDNA library."
BASE COUNT      113 a    193 c    175 g    159 t
ORIGIN
```

```
  Query Match           11.9%;  Score 207.4;  DB 13;  Length 640;
  Best Local Similarity  59.1%;  Pred. No. 5.4e-47;
  Matches  374;  Conservative    0;  Mismatches  256;  Indels    3;  Gaps    1;

Qy    769 TACTTTCAGAGGGTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTG 828
          ||||| ||| |||| || ||        | ||       || ||  | |||||||   |
Db      4 TACTTCCAGCGGGTGCTGTCCAGCAAAACGGCAGGAAGGGCCCAGCTTCTGTCCTATGTT 63

Qy    829 GCAGCTTTCGGGTGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCA 888
          |||||   ||| ||| |  ||||||||||| ||  |  | ||||| || || ||||
Db     64 GCAGCCGCCGGATGCATTTTGATGGCCATTCCCCCGGTGCTCATCGGAGCGATTGCCAAG 123

Qy    889 TCAACAGACTGGAACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCA 948
          | |||   |||||| |||| |  ||   |   ||  |||||  |||    || ||| |
Db    124 GCTACACCTTGGAACGAGACAGATTACAAGGGACCCTATCCCCTGACCGTGGACGAGACG 183

Qy    949 GACATGATTTTACCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGT 1008
          |||||||  | || || || ||||||||| |||   ||||  |   | || ||||||||
Db    184 AGCATGATTCTGCCCATGGTGCTGCAGTACCTCACGCCTGACTTCGTGTCCTTCTTTGGA 243

Qy   1009 CTTGGTGCAGTTTCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGT 1068
          | || || ||||| || || || ||||| || || || || ||  |  | || ||
Db    244 TTGGGCGCTGTTTCCGCCGCCGTGATGTCCTCCGCCGACTCCTCGGTGCTCTCCGCCGCC 303

Qy   1069 TCCATGTTTGCACGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAA 1128
          |||||||| || |||||| | ||| |  |     ||| | || || || || || | |||
Db    304 TCCATGTTCGCTCGGAACGTGTACAAATTGATTTTCCGTCAGAAGGCGTCCGAGATGGAA 363

Qy   1129 ATCGTTTGGGTTATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCC 1188
          ||| ||||||| ||||||  || |  |  ||||| | ||        || ||   |||||||
Db    364 ATCATTTGGGTGATGCGAGTCGCCATCATTGTGGTGGGCATCCTGGCTACCATCATGGCC 423

Qy   1189 TTGCTGACGAAAACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTT 1248
            |    |       |  | || ||  | |||| | |  | || || || || ||| || ||
Db    424 CTCACCATTCCCTCCATCTACGGTTTGTGGTCCATGTGCTCGGATCTGGTCTACGTCATT 483

Qy   1249 ATCTTCCCCCAG---CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCC 1305
            | ||||| |||   ||| |    || | |||    |||     ||||||| || ||   |
Db    484 CTGTTCCCGCAGCTACTGATGGTGGTGCACTTCAAGAAGCACTGCAACACGTACGGCAGC 543

Qy   1306 GTGGCAGGTTATGTTTCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTAT 1365
            || | |  ||  ||   | |||   | |  || |  | || || ||| |     ||
Db    544 CTGTCGGCATACATTGTGGCCCTGGCCATCCGACTGTCGGGCGGTGAGGCCATCTTGGGA 603

Qy   1366 CTTCAGCCCTTGATCTTCTACCCTGGCTATTAC 1398
          ||    || ||||||   || ||  ||||  ||
Db    604 CTGGCTCCATTGATCAAGTATCCCCGCTACGAC 636
```

```
RESULT 10
BM629925
LOCUS       BM629925                658 bp    mRNA    linear   EST 26-FEB-2002
DEFINITION  17000687500663 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
            19600449651480 5', mRNA sequence.
ACCESSION   BM629925
VERSION     BM629925.1  GI:18929436
KEYWORDS    EST.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 658)
  AUTHORS   Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
            ,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
  TITLE     Celera Anopheles gambiae EST project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Holt R.A.
            Celera Genomics
            45 w. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
            Plate: NU010049UI  row: D  column: 06
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .658
                     /organism="Anopheles gambiae"
                     /strain="RSP-ST (Reduced susc. to Permethrin - std.
                     chromosome)"
                     /db_xref="taxon:7165"
                     /clone="19600449651480"
                     /clone_lib="A.Gam.ad.cDNA1"
                     /dev_stage="Adult"
                     /lab_host="DH10b"
                     /note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole
                     adult mosquitoes (mixed sex) frozen on liquid nitrogen.
                     cDNA  inserts >500 bp cloned directionally into pSport 1.
                     Not 1 site is  3'. Clones available through the Malaria
                     Research and Reference  Reagent Resource Center
                     (www.malaria.mr4.org)."
BASE COUNT      106 a    199 c    186 g    167 t
ORIGIN
```

```
  Query Match           11.8%;  Score 205.2;  DB 13;  Length 658;
  Best Local Similarity  59.1%;  Pred. No. 2.3e-46;
```

Matches 370; Conservative 0; Mismatches 253; Indels 3; Gaps 1;

```
Qy    733 CTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGGGTTCTCTCTTCT 792
Db     23 CTGTTGCTAATCTTTGGCGGCATACCGTGGCAGGTCTACTTCCAGCGTGTACTTTCCAGC 82

Qy    793 TCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGGTGCCTGGTGATG 852
Db     83 AAAACGGCCGGCCGCGCTCAGATCCTTTCTTATGTGGCTGCGTTCGGCTGCATCCTGATG 142

Qy    853 GCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGGAACCAGACTGCA 912
Db    143 GCCATACCGCCCGTACTGATTGGTGCCATCGCCAAGGCAACACCTTGGAACGAGACGGAC 202

Qy    913 TATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTACCAATTGTTCTG 972
Db    203 TACAAGGGCCCGTGGCCGCTGACGACGCAGGAAACGAGCATGATCCTGCCGATGGTGCTC 262

Qy    973 CAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTTTCTGCTGCTGTT 1032
Db    263 CAGTACCTGACGCCCGACTTTGTGTCGTTCTTTGGGCTCGGGGCCGTCTCGGCCGCTGTC 322

Qy   1033 ATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCACGGAACATCTAC 1092
Db    323 ATGTCGTCGGCCGACTCCTCGGTGCTGTCCGCGTCCTCCATGTTTGCCCGCAACGTTTAC 382

Qy   1093 CAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTTATGCGAATCACA 1152
Db    383 CGGCTGATCTTCCGGCAGCGTGCCTCGGAGATGGAAATTATCTGGGTCATGCGTGTGTCG 442

Qy   1153 GTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAAACTGTGTATGGG 1212
Db    443 ATTCTGGTCGTCGGCATCCTGTCCACCATTATGGCGCTCACCATACCCTCGATCTACGGC 502

Qy   1213 CTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAGC---TGCTTTGT 1269
Db    503 TTGTGGTCAATGTGTTCCGACTTGGTGTACTGTATTCTCTTCCCGCAGCTTTTGATGGTG 562

Qy   1270 GTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTTTCTGGCCTC 1329
Db    563 GTACACTTTAAGCACCATTGCAACACATACGGCAGCTTGGCGGCGTACATCATCGCGTTC 622

Qy   1330 TTCCTGAGAATAACTGGAGGGGAGCC 1355
Db    623 ATGGTGCGACTGTCGGGCGGCGAGGC 648
```

RESULT 11
BB626456
LOCUS BB626456 652 bp mRNA linear EST 26-OCT-2001
DEFINITION BB626456 RIKEN full-length enriched, adult male diencephalon Mus musculus cDNA clone 9330188K24 5', mRNA sequence.
ACCESSION BB626456
VERSION BB626456.1 GI:16464433
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 652)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .652
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="9330188K24"
/clone_lib="RIKEN full-length enriched, adult male diencephalon"
/sex="male"
/tissue_type="diencephalon"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 140 a 193 c 168 g 148 t 3 others
ORIGIN

Query Match 11.7%; Score 203.2; DB 10; Length 652;
Best Local Similarity 86.1%; Pred. No. 8.3e-46;
Matches 223; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

```
Qy      1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60
Db    394 ATGTCTTTCCACGTAGAAGGACTGGTAGCTATTATCCTCTTCTACCTCCTTATATTTCTG 453

Qy     61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120
```

```
Db    454 GTTGGAATATGGGCTGCATGGAAAACCAAAAACAGCGGCAACCCAGAAGAGCGCAGTGAA 513

Qy    121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180

Db    514 GCCATCATAGTCGGGGGCCGTGACATTGGTTTGTTGGTTGGTGGNTTTACCATGACAGCC 573

Qy    181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240

Db    574 ACCTGNGTTGGAGGAGGCTACATCAATGGGACAGCAGAAGCAGTGTATGGGCCAGGTTGT 633

Qy    241 GGCCTAGCTTGGGCTCAGG 259

Db    634 GGTCTAGCTTGGGCTCANG 652
```

```
RESULT 12
BJ122485
LOCUS       BJ122485                624 bp    mRNA    linear   EST 23-JAN-2002
DEFINITION  BJ122485 unpublished oligo-capped cDNA library, C. elegans L1 stage
            Caenorhabditis elegans cDNA clone yk1288e07 5', mRNA sequence.
ACCESSION   BJ122485
VERSION     BJ122485.1  GI:18282624
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
            ; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 624)
  AUTHORS   Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
            and Sugano,S.
  TITLE     A complementary view of the C.elegans genome
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .624
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk1288e07"
                     /clone_lib="unpublished oligo-capped cDNA library, C.
                     elegans L1 stage"
                     /sex="hermaphrodite"
                     /tissue_type="whole animal"
                     /dev_stage="L1"
                     /note="The AD-wrmcDNA library was generated with poly(A)+
                     RNA isolated from both hermaphrodite and male N2 worms of
                     all larval stages, embryos, adults and dauers and the
                     subsequent generation of cDNAs by poly(A) priming. The
                     cDNAs were cloned into pPC86"
BASE COUNT      153 a    129 c    162 g    180 t
ORIGIN
```

```
  Query Match           10.7%;  Score 186.4;  DB 13;  Length 624;
  Best Local Similarity  60.2%;  Pred. No. 4.3e-41;
  Matches  353;  Conservative    0;  Mismatches  216;  Indels   17;  Gaps    2;

Qy      19 GGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTGGTTGGAATATGGGCTGCC 78

Db      45 GGTATCGTGGCCATTGTGTTCTTCTACGTGCTCATTCTTGTCGTTGGAATATGGGCGGGT 104

Qy      79 TGGAGAACCAAAAACAGTGGCAGC----------GCAGAAGAGCGCAGCGAAGCCATCA 127

Db     105 AGAAAATCGAAAAGTTCAAAGAGCTTGAATCAGAAGCCGGCGCGGCGACGGAAGAGGTGA 164

Qy     128 TAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCTACCTGGG 187

Db     165 TGTTAGCTGGGAGAAACATCGGAACTCTTGTCGGAATTTTCACAATGACTGCCACGTGGG 224

Qy     188 TCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTATGGCCTAG 247

Db     225 TTGGCGGTGCTTATATCAATGGAACCGCCGAGGCTCTGTATAATGGAGGT------CTCC 278

Qy     248 CTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTGTTCTTTG 307

Db     279 TTGGATGTCAGGCTCCAGTTGGATATGCAATTTCCCTTGTTATGGGAGGACTACTTTTCG 338

Qy     308 CAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAAATCTATG 367

Db     339 CAAAGAAAATGCGAGAAGAAGGATATATTACAATGCTCGATCCTTTTCAGCACAAATATG 398

Qy     368 GAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTCTGGGCTG 427

Db     399 GCCAACGAATCGGTGGCTTGATGTATGTTCCAGCACTTCTTGGTGAAACATTCTGGACAG 458

Qy     428 CAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGATATGCACA 487

Db     459 CAGCCATTCTTTCGGCACTTGGTGCAACACTGTCGGTAATTCTTGGAATCGACATGAATG 518

Qy     488 TTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGGCTCTATT 547

Db     519 CATCAGTGACCCTGTCGGCCTGTATTGCCGTATTCTACACATTCACCGGTGGATACTATG 578

Qy     548 CTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGG 593

Db     579 CAGTCGCGTACACTGACGTCGTTCAACTATTTTGCATTTTCGTCGG 624
```

```
RESULT 13
BJ125564
LOCUS       BJ125564                580 bp    mRNA    linear   EST 23-JAN-2002
DEFINITION  BJ125564 unpublished oligo-capped cDNA library, C. elegans L1 stage
            Caenorhabditis elegans cDNA clone yk1324f12 5', mRNA sequence.
ACCESSION   BJ125564
VERSION     BJ125564.1  GI:18285709
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
            ; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 580)
  AUTHORS   Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
            and Sugano,S.
  TITLE     A complementary view of the C.elegans genome
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .580
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk1324f12"
                     /clone_lib="unpublished oligo-capped cDNA library, C.
                     elegans L1 stage"
                     /sex="hermaphrodite"
                     /tissue_type="whole animal"
                     /dev_stage="L1"
                     /note="The AD-wrmcDNA library was generated with poly(A)+
                     RNA isolated from both hermaphrodite and male N2 worms of
                     all larval stages, embryos, adults and dauers and the
                     subsequent generation of cDNAs by poly(A) priming. The
                     cDNAs were cloned into pPC86"
BASE COUNT      130 a    122 c    154 g    174 t
ORIGIN
```

```
  Query Match             10.3%;  Score 179.8;  DB 13;  Length 580;
  Best Local Similarity   61.9%;  Pred. No. 3e-39;
  Matches  305;  Conservative    0;  Mismatches  182;  Indels    6;  Gaps    1;

Qy    152 TATTGGTTGGTGGATTTACCATGACAGCTACCTGGGTCGGAGGAGGGTATATCAATGGCA 211
Db      2 TGTTGGCCTACTGGTTCACAATGACTGCCACGTGGGTTGGCGGTGCTTATATCAATGGAA 61

Qy    212 CAGCTGAAGCAGTTTATGTACCAGGTTATGGCCTAGCTTGGGCTCAGGCACCAATTGGAT 271
Db     62 CCGCCGAGGCTCTGTATAATGGAGGT------CTCCTTGGATGTCAGGCTCCAGTTGGAT 115

Qy    272 ATTCTCTTAGTCTGATTTTAGGTGGCCTGTTCTTTGCAAAACCTATGCGTTCAAAGGGGT 331
Db    116 ATGCAATTTCCCTTGTTATGGGAGGACTACTTTTCGCAAAGAAAATGCGAGAAGAAGGAT 175

Qy    332 ATGTGACCATGTTAGACCCGTTTCAGCAAATCTATGGAAAACGCATGGGCGGACTCCTGT 391
Db    176 ATATTACAATGCTCGATCCTTTTCAGCACAAATATGGCCAACGAATCGGTGGCTTGATGT 235

Qy    392 TTATTCCTGCACTGATGGGAGAAATGTTCTGGGCTGCAGCAATTTTCTCTGCTTTGGGAG 451
Db    236 ATGTTCCAGCACTTCTTGGTGAAACATTCTGGACAGCAGCCATTCTTTCGGCACTTGGTG 295

Qy    452 CCACCATCAGCGTGATCATCGATGTGGATATGCACATTTCTGTCATCATCTCTGCACTCA 511
Db    296 CAACACTGTCGGTAATTCTTGGAATCGACATGAATGCATCAGTGACCCTGTCGGCCTGTA 355

Qy    512 TTGCCACTCTGTACACACTGGTGGGAGGGCTCTATTCTGTGGCCTACACTGATGTCGTTC 571
Db    356 TTGCCGTATTCTACACATTCACCGGTGGATACTATGCAGTCGCGTACACTGACGTCGTTC 415

Qy    572 AGCTCTTTTGCATTTTTGTAGGGCTGTGGATCAGCGTCCCCTTTGCATTGTCACATCCTG 631
Db    416 AACTATTTTGCATTTTCGTCGGTTTGTGGGTTTGCGTGCCGGCGGCTATGGTGCATGATG 475

Qy    632 CAGTCGCAGACAT 644
Db    476 GTGCGAAGGATAT 488


RESULT 14
AU199794
LOCUS       AU199794                 604 bp    mRNA    linear   EST 17-JUL-2001
DEFINITION  AU199794 unpublished oligo-capped cDNA library, stage L1
            Caenorhabditis elegans cDNA clone yk769e12 5', mRNA sequence.
ACCESSION   AU199794
VERSION     AU199794.1  GI:14827195
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
            ; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 604)
  AUTHORS   Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
            and Sugano,S.
  TITLE     A complementary view of the C.elegans genome
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yuji Kohara
            Genome Biology Lab.
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .604
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk769e12"
                     /clone_lib="unpublished oligo-capped cDNA library, stage
                     L1"
                     /sex="Hermaphrodite"
                     /tissue_type="whole animal"
                     /dev_stage="L1"
                     /note="The AD-wrmcDNA library was generated with poly(A)+
                     RNA isolated from both hermaphrodite and male N2 worms of
                     all larval stages, embryos, adults and dauers and the
                     subsequent generation of cDNAs by poly(A) priming. The
                     cDNAs were cloned into pPC86"
BASE COUNT      151 a    125 c    158 g    170 t
ORIGIN

  Query Match             9.5%;  Score 166;  DB 9;  Length 604;
  Best Local Similarity   58.8%;  Pred. No. 2.3e-35;
  Matches  333;  Conservative    0;  Mismatches  215;  Indels   18;  Gaps    2;

Qy     19 GGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTGGTTGGAATATGGGCTGCC 78
Db     45 GGTATCGTGGCCATTGTGTTCTTCTACGTGCTCATTCTTGTCGTTGGAATATGGGCGGGT 104

Qy     79 TGGAGAACCAAAA------------ACAGTGGCAGCGCAGAAGAGCGCAGCGAAGCCATC 126
Db    105 AGAAAATCGAAAAGTTCAAAAGAGCTTGAATCAGAAGCCGGCGCGGCGACGGAAGAGGTG 164

Qy    127 ATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCTACCTGG 186
Db    165 ATGTTAGCTGGGAGAAACATCGGAACTCTTGTCGGAATTTTCACAATGACTGCCACGTGG 224

Qy    187 GTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTATGGCCTA 246
Db    225 GTTGGCGGTGCTTATATCAATGGAACCGCCGAGGCTCTGTATAATGGAGGT------CTC 278

Qy    247 GCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTGTTCTTT 306
Db    279 CTTGGATGTCAGGCTCCAGTTGGATATGCAATTTCCCTTGTTATGGGAGGACTACTTTTC 338

Qy    307 GCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAAATCTAT 366
Db    339 GCAAAGAAAATGCGAGAAGAAGGATATATTACAATGCTCGATCCTTTTCAGCACAAATAT 398

Qy    367 GGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTCTGGGCT 426
Db    399 GGCCAACGAATCGGTGGCTTGATGTATGTTCCAGCACTTCTTGGTGAAACATTCTGGACA 458

Qy    427 GCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGATATGCAC 486
Db    459 GCAGCCATTCTTTCGGCACTTGGTGCAACACTGTCGGTAATTCTTGGAATCGACATGAAT 518

Qy    487 ATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGGCTCTAT 546
Db    519 GCATCAGTGACCCTGTCGGCCTGTATTGCCGTATTCTACACATTCACCGGTGGATACTAT 578

Qy    547 TCTGTGGCCTACACTGATGTCGTTCA 572
Db    579 GCAGTCGCGTACACTGACGTCGTTCA 604


RESULT 15
AV994375
LOCUS       AV994375                 500 bp    mRNA    linear   EST 15-MAR-2002
DEFINITION  AV994375 Nori Satoh unpublished cDNA library, tailbud embryo Ciona
            intestinalis cDNA clone citb37j13 5', mRNA sequence.
ACCESSION   AV994375
VERSION     AV994375.1  GI:19485709
KEYWORDS    EST.
SOURCE      Ciona intestinalis.
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
  TITLE     Expressed genes in Ciona intestinalis
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Nori Satoh
```

Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.

FEATURES Location/Qualifiers
source 1. .500
/organism="Ciona intestinalis"
/db_xref="taxon:7719"
/clone="citb37j13"
/clone_lib="Nori Satoh unpublished cDNA library, tailbud embryo"
/tissue_type="whole animal"
/dev_stage="tailbud embryo"
/note="Vector: pBluescript SK"

BASE COUNT 110 a 109 c 107 g 173 t 1 others
ORIGIN

Query Match 9.4%; Score 163.4; DB 10; Length 500;
Best Local Similarity 62.6%; Pred. No. 1.2e-34;
Matches 271; Conservative 0; Mismatches 161; Indels 1; Gaps 1;

```
Qy    845 TGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAG-ACTGGAAC 903
          || | ||| | || ||  | ||| | || || || ||||  ||||| |||| | ||||||
Db      1 TGTTCATGTCAATACCTTCGATATTAATCGGTGCAATTGCTGCATCTACAGNATTGGAAC 60

Qy    904 CAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTACCA 963
            ||  | || || || ||| |||       | ||  ||   || ||  ||| ||||
Db     61 GCAACATCGTACGGCCTCCCAAGTCCAGTTGAAAAAGGCGACCAAGCCAATATTCTACCC 120

Qy    964 ATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTTTCT 1023
          ||||| || || || |||  ||||||   | | || |||||||| ||||| || ||||||
Db    121 ATTGTGCTTCAATACCTCACCCCTGTAGCTGTATCATTCTTTGGGCTTGGCGCTGTTTCT 180

Qy   1024 GCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCACGG 1083
          |||||||| |||||||| || || |||   ||  | || ||   |   |||||| |  ||
Db    181 GCTGCTGTAATGTCATCTGCCGACTCTAGTATTCTATCTGCTTCTAGTATGTTTACTAGG 240

Qy   1084 AACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTTATG 1143
          || || ||  |  |   | || ||| ||| || ||||    ||  | || ||||| |||
Db    241 AATATTTATAACGTCGTCATCCGACCAAAGGCATCGGAGTTGGAGCTGGTGTGGGTAATG 300

Qy   1144 CGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAAACT 1203
           || |  || |  |  |  | |  |||   || || ||  | ||  |  |   |||| ||
Db    301 AGAGTATCAATAATATTCATAGCGGCAGGAGCCACTGCTTTAGCTCTTGTTGTGAAATCT 360

Qy   1204 GTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAGCTG 1263
          || |||   || |  || ||  |||| |||||  | ||| || |  | || || ||| |
Db    361 GTTTATTATCTTTTCTATCTTTGTTCGGACCTGATCTACGTCATACTTTTTCCACAGTTT 420

Qy   1264 CTTTGTGTACTCT 1276
           |  |||| ||||
Db    421 GTAAGTGTTCTCT 433
```

Search completed: March 31, 2003, 18:28:17
Job time : 2317 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 31, 2003, 16:26:35 ; Search time 194 Seconds
(without alignments)
7645.597 Million cell updates/sec

Title: US-10-069-541-5
Perfect score: 1743
Sequence: 1 atggctttccatgtggaagg.....ctgaagataatttacagtga 1743

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 574371 seqs, 425486471 residues

Total number of hits satisfying chosen parameters: 1148742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	242.6	13.9	1461	10	US-09-974-300-501	Sequence 501, App
	2	155	8.9	455	10	US-09-864-761-1838	Sequence 1838, Ap
	3	72	4.1	96	10	US-09-864-761-18589	Sequence 18589, A
	4	41	2.4	2839	10	US-09-995-007-1	Sequence 1, Appli
	5	39.6	2.3	2028	10	US-09-733-630-1	Sequence 1, Appli
	6	39.6	2.3	2456	10	US-09-733-630-3	Sequence 3, Appli
	7	38.2	2.2	2028	9	US-09-928-530-3	Sequence 3, Appli
	8	38.2	2.2	2028	9	US-10-162-012-28	Sequence 28, Appl
	9	38.2	2.2	2326	9	US-09-928-530-1	Sequence 1, Appli
	10	38.2	2.2	2326	9	US-10-162-012-26	Sequence 26, Appl
	11	36.6	2.1	148567	9	US-10-254-869-3	Sequence 3, Appli
	12	36.6	2.1	148567	10	US-09-801-876B-3	Sequence 3, Appli
c	13	35.6	2.0	5629	9	US-10-092-154-1162	Sequence 1162, Ap
c	14	35.6	2.0	5629	10	US-09-764-847-1162	Sequence 1162, Ap
c	15	35.6	2.0	6265	9	US-10-092-154-1161	Sequence 1161, Ap
c	16	35.6	2.0	6265	10	US-09-764-847-1161	Sequence 1161, Ap
	17	35.4	2.0	234	9	US-10-046-935-1856	Sequence 1856, Ap
	18	35.4	2.0	234	9	US-09-878-178-1856	Sequence 1856, Ap
	19	35.4	2.0	314	10	US-09-983-965-359	Sequence 359, App
c	20	35.2	2.0	173808	12	US-10-003-806-10	Sequence 10, Appl
	21	35	2.0	344	10	US-09-960-352-13002	Sequence 13002, A
	22	34.8	2.0	1776	10	US-09-815-242-9527	Sequence 9527, Ap
c	23	34.8	2.0	2712	10	US-09-974-300-1927	Sequence 1927, Ap
c	24	34.8	2.0	12482	9	US-10-270-877-25	Sequence 25, Appl
c	25	34.8	2.0	12482	9	US-10-270-837-25	Sequence 25, Appl
	26	34.2	2.0	10996	10	US-09-070-927A-202	Sequence 202, App
	27	34.2	2.0	368004	10	US-09-949-654-3	Sequence 3, Appli
	28	34	2.0	683	9	US-09-984-245-17	Sequence 17, Appl
	29	34	2.0	683	9	US-09-966-262-17	Sequence 17, Appl
c	30	34	2.0	1109	7	US-08-781-986A-222	Sequence 222, App
c	31	34	2.0	465237	10	US-09-933-267A-1	Sequence 1, Appli
	32	33.8	1.9	274	10	US-09-923-876-4803	Sequence 4803, Ap
	33	33.8	1.9	1791	9	US-10-119-988-9	Sequence 9, Appli
	34	33.8	1.9	1993	9	US-10-119-988-7	Sequence 7, Appli
c	35	33.8	1.9	3178	10	US-09-954-456-1882	Sequence 1882, Ap
c	36	33.8	1.9	3178	10	US-09-969-347-189	Sequence 189, App
	37	33.8	1.9	176373	9	US-10-095-407-17	Sequence 17, Appl
c	38	33.6	1.9	1985	10	US-09-764-870-26	Sequence 26, Appl
c	39	33.6	1.9	2857	9	US-09-978-295A-263	Sequence 263, App
c	40	33.6	1.9	2857	9	US-09-978-697-263	Sequence 263, App
c	41	33.6	1.9	2857	9	US-09-978-192A-263	Sequence 263, App
c	42	33.6	1.9	2857	9	US-10-066-500-120	Sequence 120, App
c	43	33.6	1.9	2857	9	US-09-999-832A-263	Sequence 263, App
c	44	33.6	1.9	2857	9	US-09-978-189-263	Sequence 263, App
c	45	33.6	1.9	2857	9	US-10-174-590-105	Sequence 105, App

ALIGNMENTS

RESULT 1
US-09-974-300-501
; Sequence 501, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 501
; LENGTH: 1461
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-501

Query Match 13.9%; Score 242.6; DB 10; Length 1461;
Best Local Similarity 52.4%; Pred. No. 3e-62;
Matches 663; Conservative 0; Mismatches 554; Indels 48; Gaps 4;

```
Qy    151 TTATTGGTTGGTGGATTTACCATGACAGCTACCTGGGTCGGAGGAGGGTATATCAATGGC 210
          || || || ||     | || ||| | || || |||||||| || || ||||| || ||
Db    125 TTTTTCGTCGGAATGGTGACGATGGCCGCAACATGGGTCGGCGGCGGATATATTAACGGA 184

Qy    211 ACAGCTGAAGCAGTTTATGTACCAGGTTATGGCCTAGCTTGGGCTCAGGCACCAATTGGA 270
          || || ||| |   |||        |  | ||||||    ||||| || || ||    ||
Db    185 ACGGCCGAATCGACTTACA------GCGACGGCCTCATCTGGGCCCAAGCGCCTTGGGGC 238

Qy    271 TATTCTCTTAGTCTGATTTTAGGTGGCCTGTTCTTTGCAAAACCTATGCGTTCAAAGGGG 330
          ||  |  | || |||||| | || ||  | ||||| || | |   |||||     |
Db    239 TACGCATTGAGCCTGATTATCGGCGGTATTTTCTTCGCCAGAAAAATGCGCCGTCATCAA 298

Qy    331 TATGTGACCATGTTAGACCCGTTTCAGCAAATCTATGGAAAACGCATGGGCGGACTCCTG 390
          | | |||||||  | || ||  |  | ||   ||  || || || |||| |||  |  |
```

```
Db    299 TTTATGACCATTATCGATCCTCTCGAACAGCGCTTCGGTAAGCGGATGGCCGGCGTTTTA 358

Qy    391 TTTATTCCTGCACTGATGGGAGAAATGTTCTGGGCTGCAGCAATTTTCTCTGCTTTGGGA 450
          | ||| || || ||| | |||||| |||| |||   || || || ||  | || |||||
Db    359 TATATACCGGCGCTGTTAGGAGAATTGTTTTGGAGCGCCGCGATCTTAACGGCATTGGGC 418

Qy    451 GCCACCATCAGCGTGATCATCGATGTGGATATGCACATTTCTGTCATCATCTCTGCACTC 510
            | ||  || |  ||||  |  || | ||| | || |  ||  | ||  | || ||  |
Db    419 ACGACTTTCGGAATGATTCTGAATATCGATTTCCAAACGTCGATTATTCTTTCGGCGATG 478

Qy    511 ATTGCCACTCTGTACACACTGGTGGGAGGGCTCTATTCTGTGGCCTACACTGATGTCGTT 570
          || ||||     || ||  |||  |||||  | |   | || || | ||| |||||| ||
Db    479 ATCGCCATCGCTTATACGGTGGCCGGAGGCATGTGGGCAGTTGCTTTCACAGATGTCTTT 538

Qy    571 CAGCTCTTTTGCATTTTTGTAGGGCTGTGGATCAGCGTCCCCTTTGCATTGTCACATCCT 630
          || |  ||  |||||| | |||||||   |   |||||| |||| |||||| ||
Db    539 CAAATGATTGTCATTTTGCTCGGGCTGTTTTTAGTCGTCCCATTTGTATTGTCGAATGTC 598

Qy    631 GCAG----------------------TCGCAGACATCGGGTTCACTGCTGTGCATGCCAA 668
          | |                        |  ||| ||     |   | || | ||    |
Db    599 GGTGCTCTCGATAGCGTATGGGCAAATTACAGGCACGATTTCGGCAGCAGCGCCAATCTG 658

Qy    669 ATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCT-----TGGCTT 723
          | ||       |  | ||||                ||| ||  | |     |||
Db    659 CTTCCGCCGTTGGACGGCTGGAAAAACCCGGATTGGGGAAACCTGTTTTGGAACTGGTGG 718

Qy    724 GATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGGGTT 783
          |||| |   ||  | || ||  | || || ||| |||||||  |  |||||||| | |||
Db    719 GATAATGCGCTCCTCTTAATTTTCGGAGGTATCGCATGGCAGGTGTACTTTCAGCGCGTT 778

Qy    784 CTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGGTGC 843
          || ||  |    || |  | |   ||       || |||  |  | || |   |   |||
Db    779 CTTTCGGCAAAATCGGAAAGCGCCGCCATGTGGCAGTCGATAATTGCCGGAGTGATCTGC 838

Qy    844 CTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGGAAC 903
          | |    ||||| ||   | || |||| || ||    |||     || || |||| |
Db    839 ATCATTGCCGCCATTCCGTGCGTAATCATCGGAGCTGCCGGAAACAGTACCGATTGGAGC 898

Qy    904 CAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTACCA 963
          | |   | |    |   ||| |||  |  | |                ||||||| ||
Db    899 CTGTTCGGAGCGAGCGCTCCGGATAACCCGGCG--------------ATGATTTTGCCG 943

Qy    964 ATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTTTCT 1023
            ||   |||| |    || |    ||  |   | | ||  | ||||||  |   |
Db    944 CAAACGCTTGCGTATTTGACGCCAGGAATCATCGCAGGCCTCGGCTTGGGTGCAATCGCA 1003

Qy    1024 GCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCACGG 1083
           || || || ||||||      || || || ||  | || |||   || |||   ||| ||
Db    1004 GCAGCCGTCATGTCAAGCATGGACTCATCGATTCTATCGGCATCATCAATGGCCGCATGG 1063

Qy    1084 AACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTTATG 1143
           || || ||||  |    | |||  |  || ||  |  | |||  |  |       ||  |
Db    1064 AATATTTACCGTCCGCTCATCAAGCCGAAGGCCACCCAAAAACAGCTGCAAAAAGTCGTC 1123

Qy    1144 CGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAAACT 1203
             | || |  || |||| || ||    ||| | | ||| ||  ||      ||||
Db    1124 AAACGCTCAATCATTTTGTTCGGCGCGGGAGCAGCGGTCATCGCGCTGAATGTCAAAAGC 1183

Qy    1204 GTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAGCTG 1263
           || |||    | |||||  |   ||| ||  | |||||   | || | || ||||||  |
Db    1184 GTTTATACTTTATGGTATTTGGCTTCGGATTTAGTTTATTGCATTCTTTTTCCCCAGTTA 1243

Qy    1264 CTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTTTCT 1323
              |||||| |||  || | ||    || ||| |  | || || | ||   |
Db    1244 ACAATGGCCCTCTTTTATAAAAGAGCAAATCTTTACGGGTCGATTGCTGGATTTGCAGTT 1303

Qy    1324 GGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATCTTC 1383
           |   || | ||||  |    || || || ||       |    ||| |||  |  |
Db    1304 GCAGTCATTCTGAGGCTCGGCGGTGGTGAACCCGCATTCGGCATTCCGCCGCTTCTGCCG 1363

Qy    1384 TACCC 1388
           || ||
Db    1364 TATCC 1368
```

RESULT 2
US-09-864-761-1838
; Sequence 1838, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 1838
; LENGTH: 455
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009963.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.97
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
US-09-864-761-1838

Query Match 8.9%; Score 155; DB 10; Length 455;
Best Local Similarity 100.0%; Pred. No. 3e-36;
Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    741 GATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGGGTTCTCTCTTCTTCCTCAGC 800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    266 GATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGGGTTCTCTCTTCTTCCTCAGC 325

Qy    801 CACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGGTGCCTGGTGATGGCCATCCC 860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    326 CACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGGTGCCTGGTGATGGCCATCCC 385

Qy    861 AGCCATACTCATTGGGGCCATTGGAGCATCAACAG 895
          |||||||||||||||||||||||||||||||||||
Db    386 AGCCATACTCATTGGGGCCATTGGAGCATCAACAG 420
```

RESULT 3
US-09-864-761-18589
; Sequence 18589, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 18589
; LENGTH: 96
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009963.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.97
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
; OTHER INFORMATION: NT HIT: gi|1141884, EVALUE 5.00e-33
US-09-864-761-18589

Query Match 4.1%; Score 72; DB 10; Length 96;
Best Local Similarity 100.0%; Pred. No. 9.7e-12;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    824 TCCTGGCAGCTTTCGGGTGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTG 883
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TCCTGGCAGCTTTCGGGTGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTG 60

Qy    884 GAGCATCAACAG 895
          ||||||||||||
Db     61 GAGCATCAACAG 72
```

RESULT 4
US-09-995-007-1
; Sequence 1, Application US/09995007
; Patent No. US20020123102A1
; GENERAL INFORMATION:
; APPLICANT: NANCY CARRASCO, ET AL.
; TITLE OF INVENTION: THYROID SODIUM/IODIDE SYMPORTER AND
; NUCLEIC ACID ENCODING SAME
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
; STREET: 90 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH 1.44 Mb STORAGE
; DISKETTE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/995,007
; FILING DATE: 26-No. US20020123102A1-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/595,553A
; FILING DATE: FEBRUARY 1, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: CRAIG J. ARNOLD
; REGISTRATION NUMBER: 34,287
; REFERENCE/DOCKET NUMBER: 96700/393
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2839
; TYPE: NUCLEIC ACID
; STRANDEDNESS: DOUBLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: <Unknown>
; DESCRIPTION: OLIGONUCLEOTIDE
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: RAT
; INDIVIDUAL ISOLATE: SODIUM/IODIDE SYMPORTER
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-995-007-1

Query Match 2.4%; Score 41; DB 10; Length 2839;
Best Local Similarity 49.8%; Pred. No. 0.23;
Matches 104; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

```
Qy    404 TGATGGGAGAAATGTTCTGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCG 463
          || |||     ||| | |   | |       |  ||| ||    |    || | | | |
Db    522 TGGTGGCCACGATGCTGTATACAGGCATCGTGATCTACGCGCCTGCGCTCATCCTGAACC 581

Qy    464 TGATCATCGATGTGGATATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGT 523
            | | ||   |||| ||       ||  || |   | | | | ||||   ||| |||
Db    582 AAGTGACCGGGTTGGACATCTGGGCATCGCTCCTGTCCACAGGAATCATCTGCACCTTGT 641

Qy    524 ACACACTGGTGGGAGGGCTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCA 583
          ||||    ||||| ||   |  |  | |||| ||  || |||||  | ||| |  |     |
Db    642 ACACTACCGTGGGTGGTATGAAGGCCGTGGTCTGGACAGATGTGTTCCAGGTTGTGGTAA 701

Qy    584 TTTTTGTAGGGCTGTGGATCAGCGTCCCC 612
              |  | || ||  | ||| | | | |  ||
Db    702 TGCTCGTTGGCTTCTGGGTGATCCTGGCC 730
```

RESULT 5
US-09-733-630-1
; Sequence 1, Application US/09733630
; Patent No. US20020034799A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Freidrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Abuin, Alejandro
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020034799A1el Human Transporter Protein and
; TITLE OF INVENTION: Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0106-USA
; CURRENT APPLICATION NUMBER: US/09/733,630
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: US 60/170,137
; PRIOR FILING DATE: 1999-12-10
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2028
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-733-630-1

Query Match 2.3%; Score 39.6; DB 10; Length 2028;
Best Local Similarity 45.6%; Pred. No. 0.48;
Matches 218; Conservative 0; Mismatches 254; Indels 6; Gaps 2;

```
Qy    119 AAGCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAG 178
          |||  |  | |  | | |||  | || || |  |  | |    |  ||      || ||   |
Db    167 AAGGCTACTTCCTGGCTGGAGGGGACATGGTGTGGTGGCCAGTGGGTGCATCCTTGTTTG 226

Qy    179 CTACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTT 238
          | | |   || ||| | ||  || |||| ||| |   || |   ||  |  ||   | ||
Db    227 CCAGCAATGTTGGAAGTGGACATTTCATTGGCCTGGCAGGGTCAAGTGCTGCTACGGGCA 286

Qy    239 ATGGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCC 298
            |    ||   |  ||| | | ||  | ||| |  | |  |   ||||| || |   |
Db    287 TTTCTGTA---TCAGCTTATGAACTTAATGGCTTGTTTTCTGTGCTGATGTTGGCCTGGA 343

Qy    299 TGTTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGC 358
          | ||| |  | |     ||   |  |||      ||       | |||      |||
Db    344 TCTTCCTACCCATCTACATTGCTGGTCAGGTCACCACGATGCCAGAATACCTACGGAAGC 403

Qy    359 AAATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGT 418
            ||  ||| |  | ||   |  ||||| | | |    ||  |     | || |
Db    404 GCTTCGGTGGCATCAGAATCCCCATCATCCTGGCTGTACTCTACCTATTTATCTACATCT 463
```

Qy 419 T---CTGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATG 475
| | | | | | | | ||| | | |||| | | |
Db 464 TCACCAAGATCTCGGTAGACATGTATGCAGGTGCCATCTTCATCCAGCAGTCTTTGCACC 523

Qy 476 TGGATATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGG 535
||||| || || | | | | ||| | || | |||| | | |
Db 524 TGGATCTGTACCTGGCCATAGTTGGGCTACTGGCCATCACTGCTGTATACACGGTTGCTG 583

Qy 536 GAGGGCTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGG 593
| || || | ||||| |||||| |||| | | ||| | ||| || ||||
Db 584 GTGGCCTGGCTGCTGTGATCTACACGGATGCCCTGCAGACGCTGATCATGCTTATAGG 641

RESULT 6
US-09-733-630-3
; Sequence 3, Application US/09733630
; Patent No. US20020034799A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Freidrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Abuin, Alejandro
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020034799A1el Human Transporter Protein and
; TITLE OF INVENTION: Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0106-USA
; CURRENT APPLICATION NUMBER: US/09/733,630
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: US 60/170,137
; PRIOR FILING DATE: 1999-12-10
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2456
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-733-630-3

Query Match 2.3%; Score 39.6; DB 10; Length 2456;
Best Local Similarity 45.6%; Pred. No. 0.55;
Matches 218; Conservative 0; Mismatches 254; Indels 6; Gaps 2;

```
Qy    119 AAGCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAG 178
          |||  |  | |  | | |||  | || || |  |  | |    |  ||      || ||   |
Db    480 AAGGCTACTTCCTGGCTGGAGGGGACATGGTGTGGTGGCCAGTGGGTGCATCCTTGTTTG 539

Qy    179 CTACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTT 238
          | | |   || ||| | ||  || |||| ||| |   || |   ||  |  ||   | ||
Db    540 CCAGCAATGTTGGAAGTGGACATTTCATTGGCCTGGCAGGGTCAAGTGCTGCTACGGGCA 599

Qy    239 ATGGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCC 298
            |    ||   |  ||| | | ||  | ||| |  | |  |   ||||| || |   |
Db    600 TTTCTGTA---TCAGCTTATGAACTTAATGGCTTGTTTTCTGTGCTGATGTTGGCCTGGA 656

Qy    299 TGTTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGC 358
          | ||| |  | |     ||   |  |||      ||       | |||      |||
Db    657 TCTTCCTACCCATCTACATTGCTGGTCAGGTCACCACGATGCCAGAATACCTACGGAAGC 716

Qy    359 AAATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGT 418
            ||  ||| |  | ||   |  ||||| | | |    ||  |     | || |
Db    717 GCTTCGGTGGCATCAGAATCCCCATCATCCTGGCTGTACTCTACCTATTTATCTACATCT 776

Qy    419 T---CTGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATG 475
          |   | |   | | |   | | ||| |   |  | ||||   |   | |
Db    777 TCACCAAGATCTCGGTAGACATGTATGCAGGTGCCATCTTCATCCAGCAGTCTTTGCACC 836

Qy    476 TGGATATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGG 535
          ||||| || || |  |  | |              ||| |  || | ||||| | | |
Db    837 TGGATCTGTACCTGGCCATAGTTGGGCTACTGGCCATCACTGCTGTATACACGGTTGCTG 896
```

```
Qy    536 GAGGGCTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGG 593
          | || ||   | |||||  |||||| |||| | | |||     |   |||  || ||||
Db    897 GTGGCCTGGCTGCTGTGATCTACACGGATGCCCTGCAGACGCTGATCATGCTTATAGG 954


RESULT 7
US-09-928-530-3
; Sequence 3, Application US/09928530
; Patent No. US20020156002A1
; GENERAL INFORMATION:
;  APPLICANT: Curtis, Rory A. J.
;  APPLICANT:  Silos-Santiago, Inmaculada
;  TITLE OF INVENTION: 32620, A NOVEL HUMAN SODIUM-SUGAR
;  TITLE OF INVENTION:  SYMPORTER FAMILY MEMBER AND USES THEREOF
;  FILE REFERENCE: 10446-080001
;  CURRENT APPLICATION NUMBER: US/09/928,530
;  CURRENT FILING DATE:  2001-08-13
;  PRIOR APPLICATION NUMBER: 60/227,068
;  PRIOR FILING DATE: 2000-08-22
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 2028
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-928-530-3

  Query Match             2.2%;  Score 38.2;  DB 9;  Length 2028;
  Best Local Similarity   46.1%;  Pred. No. 1.3;
  Matches  239;  Conservative    0;  Mismatches  273;  Indels    7;  Gaps    3;

Qy     79 TGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAAGCCATCATAGTTGGTGGC 138
          ||||  |      ||||||    |  | |    | ||| |||    |  |   |||  |
Db    126 TGGACTATGGTCCACAGTGAAGACCAAAAGAGACACAGTGAAAGGCTACTTCCTGGCTGA 185

Qy    139 CGAGATATTGGTTTATTGGTTGGTGGATTTA-CCATGACAGCTACCTGGGTCGGAGGAGG 197
          | ||   |||| |  |||   |||| |  | || ||   || | |   || ||| | ||
Db    186 AGGGAACATGGTGTGGTGGCCAGTGGGTGCATCCTTGTTTGCCAGCAATGTTGGAAGTGG 245

Qy    198 GTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTATGGCCTAGCTTGGGCTCA 257
          || ||| ||||   || |   ||| |  ||   | ||   |    ||   |  ||| |
Db    246 ACATTTCATTGGCCTGGCAGGGTCAGGTGCTGCTACGGGCATTTCTGTA---TCAGCTTA 302

Qy    258 GGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTGTTCTTTGCAAAACCTAT 317
          | ||  | ||| |  | |  |   ||||| || |   |  | ||| |  | |     ||
Db    303 TGAACTTAATGGCTTGTTTTCTGTGCTGATGTTGGCCTGGATCTTCCTACCCATCTACAT 362

Qy    318 GCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAAATCTATGGAAAACGCAT 377
          |    |||       ||       | |||     |||   ||  ||| |   | ||
Db    363 TGCTGGTCAGGTCACCACGATGCCAGAATACCTACGGAAGCGCTTCGGTGGCATCAGAAT 422

Qy    378 GGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTCTGGGCTGCAGCAATTTT 437
          |    |||||  | | |     ||  |     | || |||          |  |
Db    423 CCCCATCATCCTGGCTGTACTCTACCTATTTATCTACATCTTCACCAAGATCTCGGTAGA 482

Qy    438 CTCTGCTTTGGGAGCCACCATCAGCGTGATCATCG---ATGTGGATATGCACATTTCTGT 494
          |    |   || |||| | ||| |  |           |  ||||| || || |  | |
Db    483 CATGTATGCAGGTGCCATCTTCATCCAGCAGTCTTCGCACCTGGATCTGTACCTGGCCAT 542

Qy    495 CATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGGCTCTATTCTGTGGC 554
          |          |||  |  || | |||||  | |  || || ||   | |||||
Db    543 AGTTGGGCTACTGGCCATCACTGCTGTATACACGGTTGCTGGTGGCCTGGCTGCTGTGAT 602

Qy    555 CTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGG 593
          |||||| |||| | | |||     |   |||  || ||||
Db    603 CTACACGGATGCCCTGCAGACGCTGATCATGCTTATAGG 641


RESULT 8
US-10-162-012-28
; Sequence 28, Application US/10162012
; Publication No. US20030051660A1
; GENERAL INFORMATION:
;  APPLICANT: Curtis, Rory A.J.
;  APPLICANT:  Silos-Santiago, Inmaculada
;  APPLICANT:  Gu, Wei
;  TITLE OF INVENTION: NOVEL HUMAN ION CHANNEL AND TRANSPORTER FAMILY MEMBERS
;  FILE REFERENCE: 10448-190001
;  CURRENT APPLICATION NUMBER: US/10/162,012
;  CURRENT FILING DATE:  2002-06-04
;  PRIOR APPLICATION NUMBER: US 60/209,845
;  PRIOR FILING DATE: 2000-06-06
;  PRIOR APPLICATION NUMBER: US 09/875,321
;  PRIOR FILING DATE: 2001-06-06
;  PRIOR APPLICATION NUMBER: PCT/US01/18340
;  PRIOR FILING DATE: 2001-06-06
;  PRIOR APPLICATION NUMBER: US 60/209,257
;  PRIOR FILING DATE: 2000-06-05
;  PRIOR APPLICATION NUMBER: US 09/875,423
;  PRIOR FILING DATE: 2001-06-05
;  PRIOR APPLICATION NUMBER: PCT/US01/18398
;  PRIOR FILING DATE: 2001-06-05
;  PRIOR APPLICATION NUMBER: US 60/209,238
;  PRIOR FILING DATE: 2000-06-05
;  PRIOR APPLICATION NUMBER: US 09/875,363
;  PRIOR FILING DATE: 2001-06-05
;  PRIOR APPLICATION NUMBER: PCT/US01/18247
;  PRIOR FILING DATE: 2001-06-05
;  PRIOR APPLICATION NUMBER: US 60/227,068
;  PRIOR FILING DATE: 2000-08-22
;  PRIOR APPLICATION NUMBER: US 09/928,530
;  PRIOR FILING DATE: 2001-08-13
;  PRIOR APPLICATION NUMBER: PCT/US01/25475
;  PRIOR FILING DATE: 2001-08-15
;  PRIOR APPLICATION NUMBER: US 60/226,770
;  PRIOR FILING DATE: 2000-08-21
;  PRIOR APPLICATION NUMBER: US 09/934,421
;  PRIOR FILING DATE: 2001-08-21
;  PRIOR APPLICATION NUMBER: PCT/US01/26096
;  PRIOR FILING DATE: 2001-08-21
;  PRIOR APPLICATION NUMBER: US 60/279,281
;  PRIOR FILING DATE: 2001-03-28
;  PRIOR APPLICATION NUMBER: US 10/109,029
;  PRIOR FILING DATE: 2002-03-28
;  PRIOR APPLICATION NUMBER: PCT/US02/09728
;  PRIOR FILING DATE: 2002-03-28
;  PRIOR APPLICATION NUMBER: US 60/290,288
;  PRIOR FILING DATE: 2001-05-11
;  PRIOR APPLICATION NUMBER: US (not assigned)
;  PRIOR FILING DATE: 2002-05-13
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 28
;   LENGTH: 2028
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-162-012-28

  Query Match             2.2%;  Score 38.2;  DB 9;  Length 2028;
  Best Local Similarity   46.1%;  Pred. No. 1.3;
  Matches  239;  Conservative    0;  Mismatches  273;  Indels    7;  Gaps    3;

Qy     79 TGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAAGCCATCATAGTTGGTGGC 138
          ||||  |      ||||||    |  | |    | ||| |||    |  |   |||  |
Db    126 TGGACTATGGTCCACAGTGAAGACCAAAAGAGACACAGTGAAAGGCTACTTCCTGGCTGA 185

Qy    139 CGAGATATTGGTTTATTGGTTGGTGGATTTA-CCATGACAGCTACCTGGGTCGGAGGAGG 197
          | ||   |||| |  |||   |||| |  | || ||   || | |   || ||| | ||
Db    186 AGGGAACATGGTGTGGTGGCCAGTGGGTGCATCCTTGTTTGCCAGCAATGTTGGAAGTGG 245

Qy    198 GTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTATGGCCTAGCTTGGGCTCA 257
          || ||| ||||   || |   ||| |  ||   | ||   |    ||   |  ||| |
Db    246 ACATTTCATTGGCCTGGCAGGGTCAGGTGCTGCTACGGGCATTTCTGTA---TCAGCTTA 302

Qy    258 GGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTGTTCTTTGCAAAACCTAT 317
          | ||  | ||| |  | |  |   ||||| || |   |  | ||| |  | |     ||
```

```
Db    303 TGAACTTAATGGCTTGTTTTCTGTGCTGATGTTGGCCTGGATCTTCCTACCCATCTACAT 362

Qy    318 GCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAAATCTATGGAAAACGCAT 377

Db    363 TGCTGGTCAGGTCACCACGATGCCAGAATACCTACGGAAGCGCTTCGGTGGCATCAGAAT 422

Qy    378 GGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTCTGGGCTGCAGCAATTTT 437

Db    423 CCCCATCATCCTGGCTGTACTCTACCTATTTATCTACATCTTCACCAAGATCTCGGTAGA 482

Qy    438 CTCTGCTTTGGGAGCCACCATCAGCGTGATCATCG---ATGTGGATATGCACATTTCTGT 494

Db    483 CATGTATGCAGGTGCCATCTTCATCCAGCAGTCTTCGCACCTGGATCTGTACCTGGCCAT 542

Qy    495 CATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGGCTCTATTCTGTGGC 554

Db    543 AGTTGGGCTACTGGCCATCACTGCTGTATACACGGTTGCTGGTGGCCTGGCTGCTGTGAT 602

Qy    555 CTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGG 593

Db    603 CTACACGGATGCCCTGCAGACGCTGATCATGCTTATAGG 641
```

RESULT 9
US-09-928-530-1
; Sequence 1, Application US/09928530
; Patent No. US20020156002A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A. J.
; APPLICANT: Silos-Santiago, Inmaculada
; TITLE OF INVENTION: 32620, A NOVEL HUMAN SODIUM-SUGAR
; TITLE OF INVENTION: SYMPORTER FAMILY MEMBER AND USES THEREOF
; FILE REFERENCE: 10446-080001
; CURRENT APPLICATION NUMBER: US/09/928,530
; CURRENT FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/227,068
; PRIOR FILING DATE: 2000-08-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2326
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (178)...(2202)
US-09-928-530-1

Query Match 2.2%; Score 38.2; DB 9; Length 2326;
Best Local Similarity 46.1%; Pred. No. 1.4;
Matches 239; Conservative 0; Mismatches 273; Indels 7; Gaps 3;

```
Qy     79 TGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAAGCCATCATAGTTGGTGGC 138

Db    303 TGGACTATGGTCCACAGTGAAGACCAAAAGAGACACAGTGAAAGGCTACTTCCTGGCTGA 362

Qy    139 CGAGATATTGGTTTATTGGTTGGTGGATTTA-CCATGACAGCTACCTGGGTCGGAGGAGG 197

Db    363 AGGGAACATGGTGTGGTGGCCAGTGGGTGCATCCTTGTTTGCCAGCAATGTTGGAAGTGG 422

Qy    198 GTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTATGGCCTAGCTTGGGCTCA 257

Db    423 ACATTTCATTGGCCTGGCAGGGTCAGGTGCTGCTACGGGCATTTCTGTA---TCAGCTTA 479

Qy    258 GGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTGTTCTTTGCAAAACCTAT 317

Db    480 TGAACTTAATGGCTTGTTTTCTGTGCTGATGTTGGCCTGGATCTTCCTACCCATCTACAT 539

Qy    318 GCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAAATCTATGGAAAACGCAT 377

Db    540 TGCTGGTCAGGTCACCACGATGCCAGAATACCTACGGAAGCGCTTCGGTGGCATCAGAAT 599

Qy    378 GGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTCTGGGCTGCAGCAATTTT 437

Db    600 CCCCATCATCCTGGCTGTACTCTACCTATTTATCTACATCTTCACCAAGATCTCGGTAGA 659

Qy    438 CTCTGCTTTGGGAGCCACCATCAGCGTGATCATCG---ATGTGGATATGCACATTTCTGT 494

Db    660 CATGTATGCAGGTGCCATCTTCATCCAGCAGTCTTCGCACCTGGATCTGTACCTGGCCAT 719

Qy    495 CATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGGCTCTATTCTGTGGC 554

Db    720 AGTTGGGCTACTGGCCATCACTGCTGTATACACGGTTGCTGGTGGCCTGGCTGCTGTGAT 779

Qy    555 CTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGG 593

Db    780 CTACACGGATGCCCTGCAGACGCTGATCATGCTTATAGG 818
```

RESULT 10
US-10-162-012-26
; Sequence 26, Application US/10162012
; Publication No. US20030051660A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A.J.
; APPLICANT: Silos-Santiago, Inmaculada
; APPLICANT: Gu, Wei
; TITLE OF INVENTION: NOVEL HUMAN ION CHANNEL AND TRANSPORTER FAMILY MEMBERS
; FILE REFERENCE: 10448-190001
; CURRENT APPLICATION NUMBER: US/10/162,012
; CURRENT FILING DATE: 2002-06-04
; PRIOR APPLICATION NUMBER: US 60/209,845
; PRIOR FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: US 09/875,321
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: PCT/US01/18340
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/209,257
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 09/875,423
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: PCT/US01/18398
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/209,238
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 09/875,363
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: PCT/US01/18247
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/227,068
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: US 09/928,530
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/25475
; PRIOR FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: US 60/226,770
; PRIOR FILING DATE: 2000-08-21
; PRIOR APPLICATION NUMBER: US 09/934,421
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: PCT/US01/26096
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/279,281
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 10/109,029
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: PCT/US02/09728
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 60/290,288
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US (not assigned)
; PRIOR FILING DATE: 2002-05-13
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 2326
; TYPE: DNA

```
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (178)...(2202)
US-10-162-012-26

  Query Match             2.2%;  Score 38.2;  DB 9;  Length 2326;
  Best Local Similarity   46.1%;  Pred. No. 1.4;
  Matches  239;  Conservative    0;  Mismatches  273;  Indels    7;  Gaps    3;

Qy      79 TGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAAGCCATCATAGTTGGTGGC 138
           ||||  |      ||||||    | | |    | ||| |||    | |   |||  |
Db     303 TGGACTATGGTCCACAGTGAAGACCAAAAGAGACACAGTGAAAGGCTACTTCCTGGCTGA 362

Qy     139 CGAGATATTGGTTTATTGGTTGGTGGATTTA-CCATGACAGCTACCTGGGTCGGAGGAGG 197
           | ||   |||| |  |||   |||| |  | || ||   || | |   || ||| | ||
Db     363 AGGGAACATGGTGTGGTGGCCAGTGGGTGCATCCTTGTTTGCCAGCAATGTTGGAAGTGG 422

Qy     198 GTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTATGGCCTAGCTTGGGCTCA 257
           || ||| ||||   || |   ||| |  ||   | ||   |    ||   |  ||| |
Db     423 ACATTTCATTGGCCTGGCAGGGTCAGGTGCTGCTACGGGCATTTCTGTA---TCAGCTTA 479

Qy     258 GGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTGTTCTTTGCAAAACCTAT 317
           | ||  | ||| |  | | |   ||||| || |   |  | ||| |  | |     ||
Db     480 TGAACTTAATGGCTTGTTTTCTGTGCTGATGTTGGCCTGGATCTTCCTACCCATCTACAT 539

Qy     318 GCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAAATCTATGGAAAACGCAT 377
           |    |||        ||        | |||      |||   ||  ||| |   | ||
Db     540 TGCTGGTCAGGTCACCACGATGCCAGAATACCTACGGAAGCGCTTCGGTGGCATCAGAAT 599

Qy     378 GGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTCTGGGCTGCAGCAATTTT 437
           |    |||||  | | |     ||  |     | || |||                |  |
Db     600 CCCCATCATCCTGGCTGTACTCTACCTATTTATCTACATCTTCACCAAGATCTCGGTAGA 659

Qy     438 CTCTGCTTTGGGAGCCACCATCAGCGTGATCATCG---ATGTGGATATGCACATTTCTGT 494
           |      |   || |||| | ||| |  |          |  ||||| || || |  | |
Db     660 CATGTATGCAGGTGCCATCTTCATCCAGCAGTCTTCGCACCTGGATCTGTACCTGGCCAT 719

Qy     495 CATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGGCTCTATTCTGTGGC 554
           |           |||  |  || | |||||  | |  || || ||   | |||||
Db     720 AGTTGGGCTACTGGCCATCACTGCTGTATACACGGTTGCTGGTGGCCTGGCTGCTGTGAT 779

Qy     555 CTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGG 593
           |||||| |||| | | |||    |   |||  || ||||
Db     780 CTACACGGATGCCCTGCAGACGCTGATCATGCTTATAGG 818


RESULT 11
US-10-254-869-3
; Sequence 3, Application US/10254869
; Publication No. US20030027307A1
; GENERAL INFORMATION:
;  APPLICANT: YE, Jane et al
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001160DIV
;  CURRENT APPLICATION NUMBER: US/10/254,869
;  CURRENT FILING DATE:  2002-09-26
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 148567
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(148567)
;   OTHER INFORMATION: n = A,T,C or G
US-10-254-869-3

  Query Match             2.1%;  Score 36.6;  DB 9;  Length 148567;
  Best Local Similarity   46.1%;  Pred. No. 80;
  Matches  123;  Conservative    0;  Mismatches  144;  Indels    0;  Gaps    0;

Qy    1416 ATATAATCAGAAATTTCCATTTAAAACACTTGCCATGGTTACATCATTCTTAACCAACAT 1475
           | ||||||||   |||  | ||  ||      ||  | |    || | | ||    ||| |
Db   31673 AGATAATCAGTTGTTTTAACTTTTAATTTAAGCAGTAGCAGAATGACTTTTTGGGAACTT 31732

Qy    1476 TTGCATCTCCTATCTAGCCAAGTATCTATTTGAAAGTGGAACCTTGCCACCTAAATTAGA 1535
           | || |   | |       | | | | | |  ||    |||  ||  |  ||   ||
Db   31733 AGGAATTTGGAAACCTTTTTATTCTATGTATTGAATATCAACTATGTAATTTAGTCTAAG 31792

Qy    1536 TGTATTTGATGCTGTTGTTGCAAGACACAGTGAAGAAAACATGGATAAGACAATTCTTGT 1595
           |||  || |         || ||| |   |  | || |   |||       | ||| | |||
Db   31793 GTTATATGCTAGAAACATTTCAAAAACGAAAGCAGCAGCAATGACATCAAAAATGCATGT 31852

Qy    1596 CAAAAATGAAAATATTAAATTAGATGAACTTGCACTTGTGAAGCCACGACAGAGCATGAC 1655
           |||||   ||    || || ||||   || |    ||   ||  |     |      | ||
Db   31853 CAAAAGCAAATGGTTTTAAATAGAAATACATCATTTTAACAATCTTGAAGTTTAAAAGAT 31912

Qy    1656 CCTCAGCTCAACTTTCACCAATAAAGA 1682
           |||      | |    ||| | || ||
Db   31913 CCTATAAAAATCACAAACCCAGAAGGA 31939


RESULT 12
US-09-801-876B-3
; Sequence 3, Application US/09801876B
; Patent No. US20020127683A1
; GENERAL INFORMATION:
;  APPLICANT: YE, Jane et al
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001160
;  CURRENT APPLICATION NUMBER: US/09/801,876B
;  CURRENT FILING DATE:  2001-03-09
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 148567
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(148567)
;   OTHER INFORMATION: n = A,T,C or G
US-09-801-876B-3

  Query Match             2.1%;  Score 36.6;  DB 10;  Length 148567;
  Best Local Similarity   46.1%;  Pred. No. 80;
  Matches  123;  Conservative    0;  Mismatches  144;  Indels    0;  Gaps    0;

Qy    1416 ATATAATCAGAAATTTCCATTTAAAACACTTGCCATGGTTACATCATTCTTAACCAACAT 1475
           | ||||||||   |||  | ||  ||      ||  | |    || | | ||    ||| |
Db   31673 AGATAATCAGTTGTTTTAACTTTTAATTTAAGCAGTAGCAGAATGACTTTTTGGGAACTT 31732

Qy    1476 TTGCATCTCCTATCTAGCCAAGTATCTATTTGAAAGTGGAACCTTGCCACCTAAATTAGA 1535
           | || |   | |       | | | | | |  ||    |||  ||  |  ||   ||
Db   31733 AGGAATTTGGAAACCTTTTTATTCTATGTATTGAATATCAACTATGTAATTTAGTCTAAG 31792

Qy    1536 TGTATTTGATGCTGTTGTTGCAAGACACAGTGAAGAAAACATGGATAAGACAATTCTTGT 1595
           |||  || |         || ||| |   |  | || |   |||       | ||| | |||
Db   31793 GTTATATGCTAGAAACATTTCAAAAACGAAAGCAGCAGCAATGACATCAAAAATGCATGT 31852

Qy    1596 CAAAAATGAAAATATTAAATTAGATGAACTTGCACTTGTGAAGCCACGACAGAGCATGAC 1655
           |||||   ||    || || ||||   || |    ||   ||  |     |      | ||
Db   31853 CAAAAGCAAATGGTTTTAAATAGAAATACATCATTTTAACAATCTTGAAGTTTAAAAGAT 31912

Qy    1656 CCTCAGCTCAACTTTCACCAATAAAGA 1682
           |||      | |    ||| | || ||
Db   31913 CCTATAAAAATCACAAACCCAGAAGGA 31939


RESULT 13
```

US-10-092-154-1162/c
; Sequence 1162, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1162
; LENGTH: 5629
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-154-1162

Query Match 2.0%; Score 35.6; DB 9; Length 5629;
Best Local Similarity 51.5%; Pred. No. 16;
Matches 106; Conservative 0; Mismatches 99; Indels 1; Gaps 1;

```
Qy    743 TGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGGGTTCTCTCTTCTTCCTCAGCCA 802
          ||||| ||         |||  ||||     |||       ||| || |  | ||||||
Db   1056 TGCTGTGTCACCCAGGCTGGAGAGCAATGGCTCAATCTTGGCTCACTGCAACTTCAGCCT 997

Qy    803 CCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGGTGCCTGGTGATGGCCATCCCAG 862
          || | | |||||||  |   |||  |      ||  || ||  ||||     |||| | ||
Db    996 CCGAGGTTCAAGTGATTCTCTTGCCTCAGCCTCCCGAGTAGCTGGCACATGCCACCACA- 938

Qy    863 CCATACTCATTGGGGCCATTGGAGCATCAACAGACTGGAACCAGACTGCATATGGGCTTC 922
          ||   || |||   |   ||  || |   || |  |   |||| | ||  || |    ||
Db    937 CCCGGCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATATTGGCTAAGCTGGTC 878

Qy    923 CAGATCCCAAGACTACAGAAGAGGCA 948
          || | |  |||  |||  ||  ||
Db    877 TCGAACTCCTGACCTCAGGTGATCCA 852
```

RESULT 14
US-09-764-847-1162/c
; Sequence 1162, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1162
; LENGTH: 5629
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1162

Query Match 2.0%; Score 35.6; DB 10; Length 5629;
Best Local Similarity 51.5%; Pred. No. 16;
Matches 106; Conservative 0; Mismatches 99; Indels 1; Gaps 1;

```
Qy    743 TGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGGGTTCTCTCTTCTTCCTCAGCCA 802
          ||||| ||         |||  ||||     |||       ||| || |  | ||||||
Db   1056 TGCTGTGTCACCCAGGCTGGAGAGCAATGGCTCAATCTTGGCTCACTGCAACTTCAGCCT 997

Qy    803 CCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGGTGCCTGGTGATGGCCATCCCAG 862
          || | | |||||||  |   |||  |      ||  || ||  ||||     |||| | ||
Db    996 CCGAGGTTCAAGTGATTCTCTTGCCTCAGCCTCCCGAGTAGCTGGCACATGCCACCACA- 938

Qy    863 CCATACTCATTGGGGCCATTGGAGCATCAACAGACTGGAACCAGACTGCATATGGGCTTC 922
          ||   || |||   |   ||  || |   || |  |   |||| | ||  || |    ||
Db    937 CCCGGCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATATTGGCTAAGCTGGTC 878

Qy    923 CAGATCCCAAGACTACAGAAGAGGCA 948
          || | |  |||  |||  ||  ||
Db    877 TCGAACTCCTGACCTCAGGTGATCCA 852
```

RESULT 15
US-10-092-154-1161/c
; Sequence 1161, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1161
; LENGTH: 6265
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-154-1161

Query Match 2.0%; Score 35.6; DB 9; Length 6265;
Best Local Similarity 51.5%; Pred. No. 17;
Matches 106; Conservative 0; Mismatches 99; Indels 1; Gaps 1;

```
Qy    743 TGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGGGTTCTCTCTTCTTCCTCAGCCA 802
          ||||| ||         |||  ||||     |||       ||| || |  | ||||||
Db   1056 TGCTGTGTCACCCAGGCTGGAGAGCAATGGCTCAATCTTGGCTCACTGCAACTTCAGCCT 997

Qy    803 CCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGGTGCCTGGTGATGGCCATCCCAG 862
          || | | |||||||  |   |||  |      ||  || ||  ||||     |||| | ||
Db    996 CCGAGGTTCAAGTGATTCTCTTGCCTCAGCCTCCCGAGTAGCTGGCACATGCCACCACA- 938

Qy    863 CCATACTCATTGGGGCCATTGGAGCATCAACAGACTGGAACCAGACTGCATATGGGCTTC 922
          ||   || |||   |   ||  || |   || |  |   |||| | ||  || |    ||
Db    937 CCCGGCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATATTGGCTAAGCTGGTC 878

Qy    923 CAGATCCCAAGACTACAGAAGAGGCA 948
          || | |  |||  |||  ||  ||
Db    877 TCGAACTCCTGACCTCAGGTGATCCA 852
```

Search completed: March 31, 2003, 18:36:11
Job time : 377 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 31, 2003, 16:23:55 ; Search time 84 Seconds
(without alignments)
6363.543 Million cell updates/sec

Title: US-10-069-541-5
Perfect score: 1743
Sequence: 1 atggctttccatgtggaagg.....ctgaagataatttacagtga 1743

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	47.6	2.7	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	41	2.4	2839	4	US-08-595-553A-1	Sequence 1, Appli
c	3	36.6	2.1	4160	4	US-09-134-218-1	Sequence 1, Appli
c	4	35.6	2.0	2397	4	US-09-221-017B-272	Sequence 272, App
	5	35.2	2.0	2238	1	US-07-841-651-1	Sequence 1, Appli
	6	34.8	2.0	1593	4	US-09-134-001C-1673	Sequence 1673, Ap
c	7	34.8	2.0	25002	4	US-08-961-527-48	Sequence 48, Appl
	8	34.6	2.0	2847	4	US-09-484-970B-22	Sequence 22, Appl
	9	34.2	2.0	1515	4	US-09-071-035-431	Sequence 431, App
	10	34.2	2.0	1803	4	US-09-071-035-429	Sequence 429, App
c	11	33.8	1.9	3172	1	US-07-741-940-3	Sequence 3, Appli
c	12	33.8	1.9	3172	1	US-08-289-548A-3	Sequence 3, Appli
c	13	33.8	1.9	3172	1	US-08-452-654-3	Sequence 3, Appli
c	14	33.8	1.9	3172	1	US-08-452-655B-3	Sequence 3, Appli
c	15	33.8	1.9	3172	3	US-08-450-582-3	Sequence 3, Appli
c	16	33.8	1.9	3172	4	US-08-449-731-3	Sequence 3, Appli
	17	33.8	1.9	176373	3	US-09-128-155-17	Sequence 17, Appl
	18	33.6	1.9	84495	4	US-09-797-906-3	Sequence 3, Appli
	19	33.4	1.9	3593	4	US-09-404-627-3	Sequence 3, Appli
	20	33.4	1.9	4205	4	US-09-404-627-1	Sequence 1, Appli
c	21	33.4	1.9	48974	4	US-08-920-422-17	Sequence 17, Appl
	22	33.2	1.9	4396	4	US-09-821-736-1	Sequence 1, Appli
c	23	32.8	1.9	5232	4	US-08-972-927-1	Sequence 1, Appli
c	24	32.8	1.9	9936	4	US-08-972-927-2	Sequence 2, Appli
	25	32.6	1.9	4739	3	US-08-685-871-1	Sequence 1, Appli
	26	32.4	1.9	316	4	US-09-118-554-18	Sequence 18, Appl
	27	32.4	1.9	316	4	US-09-118-627-18	Sequence 18, Appl
	28	32.4	1.9	316	4	US-09-602-877A-18	Sequence 18, Appl
c	29	32.4	1.9	2101	1	US-08-106-761-1	Sequence 1, Appli
	30	32	1.8	783	4	US-09-134-001C-2035	Sequence 2035, Ap
	31	32	1.8	11770	4	US-08-961-527-172	Sequence 172, App
c	32	31.8	1.8	933	4	US-09-134-001C-1564	Sequence 1564, Ap
	33	31.8	1.8	1272	4	US-09-134-001C-1880	Sequence 1880, Ap
	34	31.8	1.8	3182	4	US-08-971-395-1	Sequence 1, Appli
	35	31.8	1.8	3183	1	US-08-413-135-1	Sequence 1, Appli
	36	31.8	1.8	4467	1	US-08-565-907A-1	Sequence 1, Appli
	37	31.8	1.8	4467	2	US-08-910-551B-1	Sequence 1, Appli
	38	31.8	1.8	4467	2	US-08-909-425A-1	Sequence 1, Appli
	39	31.8	1.8	35081	2	US-08-752-760A-1	Sequence 1, Appli
	40	31.6	1.8	1668	4	US-09-813-872-3	Sequence 3, Appli
c	41	31.4	1.8	1787	1	US-08-278-630A-13	Sequence 13, Appl
c	42	31.4	1.8	1980	1	US-08-278-630A-12	Sequence 12, Appl
	43	31.4	1.8	3095	6	5231168-1	Patent No. 5231168
c	44	31.4	1.8	13011	2	US-08-791-849A-14	Sequence 14, Appl
	45	31.4	1.8	65042	4	US-09-784-316-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 2.7%; Score 47.6; DB 1; Length 7218;

```
  Best Local Similarity   6.0%;  Pred. No. 0.00035;
  Matches   23;  Conservative  200;  Mismatches  159;  Indels    0;  Gaps    0;

Qy     692 CTGTTGACTCATCTGAAGTCTACTCTTGGCTTGATAGTTTTCTGTTGTTGATGCTGGGTG 751
           :: ::  ::: :::     ::: :::::  :::  :  :::::: :: ::  : ::   :
Db    1088 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1147

Qy     752 GAATCCCATGGCAAGCATACTTTCAGAGGGTTCTCTCTTCTTCCTCAGCCACCTATGCTC 811
           :::: :  :   : : :::::      ::::::::::::::::  :: ::: : :::
Db    1148 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1207

Qy     812 AAGTGCTGTCCTTCCTGGCAGCTTTCGGGTGCCTGGTGATGGCCATCCCAGCCATACTCA 871
           : :: ::::::::  :  :::::   : :::  :  :  :: ::::  :: : :::
Db    1208 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1267

Qy     872 TTGGGGCCATTGGAGCATCAACAGACTGGAACCAGACTGCATATGGGCTTCCAGATCCCA 931
           ::    :: ::    : ::  :   ::    ::   :: : : :   :::::   ::::
Db    1268 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1327

Qy     932 AGACTACAGAAGAGGCAGACATGATTTTACCAATTGTTCTGCAGTATCTCTGCCCTGTGT 991
           :: :       :   : :  :::: ::  :: :::: :   : ::::: :::: : :
Db    1328 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1387

Qy     992 ATATTTCTTTCTTTGGTCTTGGTGCAGTTTCTGCTGCTGTTATGTCATCAGCAGATTCTT 1051
           : ::::::::::: ::::  : :  ::::: :: :: :: : ::   | || ||||||
Db    1388 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCAAATTCTT 1447

Qy    1052 CCATCTTGTCAGCAAGTTCCAT 1073
           | ||||  | | | | || |||
Db    1448 CTATCTCTTTAACTACTTGCAT 1469


RESULT 2
US-08-595-553A-1
; Sequence 1, Application US/08595553A
; Patent No. 6391579
;  GENERAL INFORMATION:
;    APPLICANT:  NANCY CARRASCO, ET AL.
;    TITLE OF INVENTION:  THYROID SODIUM/IODIDE SYMPORTER AND
;    TITLE OF INVENTION:  NUCLEIC ACID ENCODING SAME
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  AMSTER, ROTHSTEIN & EBENSTEIN
;      STREET:  90 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  U.S.A.
;      ZIP:  10016
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH 1.44 Mb STORAGE
;      MEDIUM TYPE:  DISKETTE
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/595,553A
;      FILING DATE:  FEBRUARY 1, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CRAIG J. ARNOLD
;      REGISTRATION NUMBER:  34,287
;      REFERENCE/DOCKET NUMBER:  96700/393
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 697-5995
;      TELEFAX:  (212) 286-0854 or 286-0082
;      TELEX:  TWX 710-581-4766
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2839
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:
;      DESCRIPTION:  OLIGONUCLEOTIDE
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  RAT
;      INDIVIDUAL ISOLATE:  SODIUM/IODIDE SYMPORTER
US-08-595-553A-1

  Query Match             2.4%;  Score 41;  DB 4;  Length 2839;
  Best Local Similarity   49.8%;  Pred. No. 0.022;
  Matches  104;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy     404 TGATGGGAGAAATGTTCTGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCG 463
           || |||      ||| | |   | |       |  |||  ||    |    || | | | |
Db     522 TGGTGGCCACGATGCTGTATACAGGCATCGTGATCTACGCGCCTGCGCTCATCCTGAACC 581

Qy     464 TGATCATCGATGTGGATATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGT 523
           | | ||  |||| ||       ||  || |   | | | | ||||   |||  |||
Db     582 AAGTGACCGGGTTGGACATCTGGGCATCGCTCCTGTCCACAGGAATCATCTGCACCTTGT 641

Qy     524 ACACACTGGTGGGAGGGCTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCA 583
           ||||    ||||| ||  |  |  | |||| ||  || |||||  | ||| |  |      |
Db     642 ACACTACCGTGGGTGGTATGAAGGCCGTGGTCTGGACAGATGTGTTCCAGGTTGTGGTAA 701

Qy     584 TTTTTGTAGGGCTGTGGATCAGCGTCCCC 612
           |  | || ||  | ||| | | | |  ||
Db     702 TGCTCGTTGGCTTCTGGGTGATCCTGGCC 730


RESULT 3
US-09-134-218-1/c
; Sequence 1, Application US/09134218A
; Patent No. 6312926
; GENERAL INFORMATION:
;  APPLICANT: Shatkin, Aaron J.
;  APPLICANT: Pillutla, Renuka
;  APPLICANT: Reinberg, Danny
;  APPLICANT: Yu, Zheng
;  APPLICANT: Moldanado, Edio
;  TITLE OF INVENTION: mRNA CAPPING ENZYMES AND USES THEREOF
;  FILE REFERENCE: 601-1-079 ss
;  CURRENT APPLICATION NUMBER: US/09/134,218A
;  CURRENT FILING DATE: 1998-08-14
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 4160
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-134-218-1

  Query Match             2.1%;  Score 36.6;  DB 4;  Length 4160;
  Best Local Similarity   46.6%;  Pred. No. 0.66;
  Matches  117;  Conservative    0;  Mismatches  134;  Indels    0;  Gaps    0;

Qy    1417 TATAATCAGAAATTTCCATTTAAAACACTTGCCATGGTTACATCATTCTTAACCAACATT 1476
           || | | | ||     |||||||||| |||    ||| ||||| | |       | |  |
Db    2323 TAAACTTAAAATCCAACATTTAAAAAACTCAATATGCTTACAGCTTCAGATTGCTAGTTA 2264

Qy    1477 TGCATCTCCTATCTAGCCAAGTATCTATTTGAAAGTGGAACCTTGCCACCTAAATTAGAT 1536
           || |||   | |  ||  |  |||  |  |  ||   |   ||       ||| ||  ||
Db    2263 TGAATCAAATGTAAAGGTATCTATTCACATACAAACAGGCTCTATTTCATTAACTTCAAT 2204

Qy    1537 GTATTTGATGCTGTTGTTGCAAGACACAGTGAAGAAAACATGGATAAGACAATTCTTGTC 1596
           |  || |  || | | |     |  | ||  |    ||| ||  | |   | | |  ||
Db    2203 CTGATTTAACCTTTGGGTATTTCAATCTGTAGACTCCACAGGGTAAGGCTGAATTTATTC 2144

Qy    1597 AAAAATGAAAATATTAAATTAGATGAACTTGCACTTGTGAAGCCACGACAGAGCATGACC 1656
           |   || || | | ||  |  |   |   || |||   |   | | || |    |||  |
Db    2143 AGGTATAAATAAAATATTTAGGTCCATGATGTACTGTAGTTCCAAGGAAACCAAATGTAC 2084

Qy    1657 CTCAGCTCAAC 1667
           |  |  |  ||
Db    2083 CAAATATATAC 2073
```

```
RESULT 4
US-09-221-017B-272/c
; Sequence 272, Application US/09221017B
; Patent No. 6444799
;  GENERAL INFORMATION:
;    APPLICANT:  Ross, Bruce C.
;    TITLE OF INVENTION:  P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
;    NUMBER OF SEQUENCES:  1120
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/221,017B
;      FILING DATE:  23-DEC-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1182
;      FILING DATE:  31-DEC-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1546
;      FILING DATE:  30-JAN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP2911
;      FILING DATE:  09-APR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU98/01023
;      FILING DATE:  10-DEC-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monroy, Gladys H
;      REGISTRATION NUMBER:  32,430
;      REFERENCE/DOCKET NUMBER:  27340-20021.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-813-5600
;      TELEFAX:  650-494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  272:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2397 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  UNKNOWN
;    ORIGINAL SOURCE:
;      ORGANISM:  PORYPHYROMONAS GINGIVALIS
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1...2397
US-09-221-017B-272

  Query Match             2.0%;  Score 35.6;  DB 4;  Length 2397;
  Best Local Similarity   67.6%;  Pred. No. 0.97;
  Matches   50;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy    1311 AGGTTATGTTTCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCA 1370
           || |||| ||||||| |||   |    |||||  || | |||||| ||| || |||  ||
Db    1184 AGTTTATTTTTCTGGACTCCCTCATTCAATAATAGGTGAGGAGCCCTATATGCATCAACA 1125

Qy    1371 GCCCTTGATCTTCT 1384
             | | |||||||
Db    1124 TATCCTAATCTTCT 1111


RESULT 5
US-07-841-651-1
; Sequence 1, Application US/07841651
; Patent No. 5410031
;  GENERAL INFORMATION:
;    APPLICANT:  Pajor, Ana M
;    APPLICANT:  Wright, Ernest M
;    TITLE OF INVENTION:  Cloning and Functional Expression of a
;    TITLE OF INVENTION:  Mammalian Na+/Nucleoside Cotransporter: A Member of the
;    TITLE OF INVENTION:  SGLT Family
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheldon & Mak
;      STREET:  225 South Lake Avenue, Ninth Floor
;      CITY:  Pasadena
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  91101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/841,651
;      FILING DATE:  19920224
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandel, SaraLynn
;      REGISTRATION NUMBER:  31,853
;      REFERENCE/DOCKET NUMBER:  8772
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (818) 796-4000
;      TELEFAX:  (818) 795-6321
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2238 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Oryctolagus cuniculus
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  7..2022
US-07-841-651-1

  Query Match             2.0%;  Score 35.2;  DB 1;  Length 2238;
  Best Local Similarity   52.8%;  Pred. No. 1.2;
  Matches   76;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy     477 GGATATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGG 536
           | ||||  ||  ||| ||||||           |||  |||   | |||||  ||   ||
Db     522 GAATATTTACGCTTCGGTCATCGCGCTCCTGGGCATCACCATGGTTTACACCGTGACAGG 581

Qy     537 AGGGCTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCT 596
           ||||||     |  ||   ||||| ||    || |||  ||||  |||  | |  |||
Db     582 AGGGCTGGCAGCGCTGATGTACACAGACACAGTGCAGACCTTTGTCATCATCGCGGGGGC 641

Qy     597 GTGGATCAGCGTCCCCTTTGCATT 620
             |  |||  |  |   |  || ||
Db     642 CTTCATCCTCACCGGTTACGCCTT 665


RESULT 6
US-09-134-001C-1673
; Sequence 1673, Application US/09134001C
; Patent No. 6380370
```

```
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1673
;   LENGTH: 1593
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1673

  Query Match             2.0%;  Score 34.8;  DB 4;  Length 1593;
  Best Local Similarity   46.1%;  Pred. No. 1.3;
  Matches  117;  Conservative    0;  Mismatches  137;  Indels    0;  Gaps    0;

Qy     934 ACTACAGAAGAGGCAGACATGATTTTACCAATTGTTCTGCAGTATCTCTGCCCTGTGTAT 993
           | ||  ||  |  |||| |    ||  ||     || |    | | || | |  |  |
Db     931 AGTATTGACAAATCAGATAACGCTTACCCTGCACTTGTAACTTCAGTATTACCAGAATGG 990

Qy     994 ATTTCTTTCTTTGGTCTTGGTGCAGTTTCTGCTGCTGTTATGTCATCAGCAGATTCTTCC 1053
             | |  |||   |  ||  |  ||| || |||  |  ||   |||   | |   ||
Db     991 GCATTTGGCTTATTTGGTGCGGTTATTTTTGGTGCAATATTGAGCTCATTTGTTGGCTCA 1050

Qy    1054 ATCTTGTCAGCAAGTTCCATGTTTGCACGGAACATCTACCAGCTTTCCTTCAGACAAAAT 1113
            |        ||| | |  | ||  |||   |  ||||  | | |   ||  || |||||
Db    1051 TTGAATAGTACAACTACACTATTGACACTCGATTTCTATAAACCTATTTTTGGAAAAAAT 1110

Qy    1114 GCTTCGGACAAAGAAATCGTTTGGGTTATGCGAATCACAGTGTTTGTGTTTGGAGCATCT 1173
           || || ||| | || | | | ||    |  ||  |     | ||  |||||     |
Db    1111 AAATCAGATAAACATATTGCTCGAGTAGGCCATATTGCTACTGTAGTCATTGGAGTTATT 1170

Qy    1174 GCAACAGCCATGGC 1187
           |    |||  | ||
Db    1171 GTTGTAGCACTTGC 1184


RESULT 7
US-08-961-527-48/c
; Sequence 48, Application US/08961527
; Patent No. 6420135
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  391
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,527
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  48:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  25002 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-527-48

  Query Match             2.0%;  Score 34.8;  DB 4;  Length 25002;
  Best Local Similarity   47.3%;  Pred. No. 7.4;
  Matches  105;  Conservative    0;  Mismatches  117;  Indels    0;  Gaps    0;

Qy    1382 TCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAAA 1441
           || ||| || ||  |    || |||  | ||  |    ||  | ||||    |  || ||
Db    5078 TCAACCTTGTCTTCTCAAATGCTGAGTACGGCTTCATCAAGAACAAATACGAAGATACAA 5019

Qy    1442 CACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTATC 1501
              | || ||      |  ||| || |||  |  |  | |  |  | ||  |   ||
Db    5018 ACAAACACTTGTTTGGTGTAGACTTCACAAACGCTGACTACGCTAAAATTGCGGAAGCTC 4959

Qy    1502 TATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGAC 1561
              |   |    ||||  |               | || |   | ||||| ||||||||  |
Db    4958 AAGGAGCTGTTGGATTCACAGTTGACCGTATCGAAGACATCGATGCAGTTGTTGCAGAAG 4899

Qy    1562 ACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATG 1603
             | ||   ||||  ||    | || | ||||||   |||
Db    4898 CTGTTAAATTGAACAAAGAAGGTAAAACTGTTGTCATCGATG 4857


RESULT 8
US-09-484-970B-22
; Sequence 22, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
;  APPLICANT:  Jones, Karen A.
;  APPLICANT:  Volkmuth, Wayne
;  APPLICANT:  Walker, Michael G.
;  TITLE OF INVENTION:  BONE REMODELING GENES
;  FILE REFERENCE:  PB-0014 US
;  CURRENT APPLICATION NUMBER: US/09/484,970B
;  CURRENT FILING DATE:  2000-01-18
;  NUMBER OF SEQ ID NOS:  172
;  SOFTWARE:  PERL Program
; SEQ ID NO 22
;   LENGTH: 2847
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6426186 215642.2CB1
;   NAME/KEY: unsure
;   LOCATION: 364-381, 572-597, 1191-1234
;   OTHER INFORMATION: a, t, c, g, or other
US-09-484-970B-22

  Query Match             2.0%;  Score 34.6;  DB 4;  Length 2847;
  Best Local Similarity   48.3%;  Pred. No. 2.2;
  Matches   97;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy    1491 AGCCAAGTATCTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGT 1550
           | ||||||| ||  |  |   ||||   |||   | | ||    |||  |  || | | |
Db    2031 ATCCAAGTAACTGGTGAACTTTGGAGGTTTGGAGCTTGAAGAGAATGGCTAAGAAGATTT 2090

Qy    1551 TGTTGCAAGACACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATAT 1610
             |   ||  |  |   ||||| |||  || | |   || |  | | ||  | |||
Db    2091 GAATTATAGGGAGGGAACAGAAATCATACATGAAAAGGTTTTACTGAGAAGGGGAAAACC 2150

Qy    1611 TAAATTAGATGAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTT 1670
           | |  |||| | || ||        ||  ||    | |   |||  |  |  || | ||
```

```
Db   2151 TTAGATAGAGGGACATGTGAAACAAAATTCATTTGAAATTTTGATTCAGACATCCATTTC 2210

Qy   1671 CACCAATAAAGAGGCCTTCCT 1691
          ||     ||| ||     |||
Db   2211 CAGTGGCAAACAGCAAAGCCT 2231


RESULT 9
US-09-071-035-431
; Sequence 431, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,035
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. Anders Brookes
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB369P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  431:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1515 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-09-071-035-431

  Query Match             2.0%;  Score 34.2;  DB 4;  Length 1515;
  Best Local Similarity   54.3%;  Pred. No. 2;
  Matches   69;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy   1506 TGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGACACAG 1565
          | ||||   ||  | ||||||||||| | ||||    || || | |||  |||       |
Db   1383 TTAAAGATAAAAATGGCCACCTAAAGTGGATGATGATGCTGTTATTGAAGCAGCCGCGTG 1442

Qy   1566 TGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGATGAACT 1625
           |||     |||||||||  | |  |   |  | || |||  | |      || ||  |
Db   1443 GGAAAGTGGCATGGATAACGCTACACGGTTTGACAAAGAAGGTGTGGGCAAAGGCGACGT 1502

Qy   1626 TGCACTT 1632
          || | ||
Db   1503 TGGAGTT 1509


RESULT 10
US-09-071-035-429
; Sequence 429, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,035
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. Anders Brookes
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB369P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  429:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1803 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-09-071-035-429

  Query Match             2.0%;  Score 34.2;  DB 4;  Length 1803;
  Best Local Similarity   54.3%;  Pred. No. 2.2;
  Matches   69;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy   1506 TGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGACACAG 1565
          | ||||   ||  | ||||||||||| | ||||    || || | |||  |||       |
Db   1643 TTAAAGATAAAAATGGCCACCTAAAGTGGATGATGATGCTGTTATTGAAGCAGCCGCGTG 1702

Qy   1566 TGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGATGAACT 1625
           |||     |||||||||  | |  |   |  | || |||  | |      || ||  |
Db   1703 GGAAAGTGGCATGGATAACGCTACACGGTTTGACAAAGAAGGTGTGGGCAAAGGCGACGT 1762

Qy   1626 TGCACTT 1632
          || | ||
Db   1763 TGGAGTT 1769


RESULT 11
US-07-741-940-3/c
; Sequence 3, Application US/07741940
; Patent No. 5352775
;  GENERAL INFORMATION:
;    APPLICANT:  ALBERTSEN, HANS
;    APPLICANT:  ANAND, RAKESH
;    APPLICANT:  CARLSON, MARY
;    APPLICANT:  GRODEN, JOANNA
;    APPLICANT:  HEDGE, PHILIP J.
;    APPLICANT:  JOSLYN, GEOFF
;    APPLICANT:  KINZLER, KENNETH
;    APPLICANT:  MARKHAM, ALEXANDER F.
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  THLIVERIS, ANDREW
;    TITLE OF INVENTION:  INHERITED AND SOMATIC MUTATIONS OF APC
;    TITLE OF INVENTION:  GENE IN COLORECTAL CANCER IN HUMANS
;    NUMBER OF SEQUENCES:  94
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner, Birch, McKie & Beckett
;      STREET:  1001 G Street, NW
```

```
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  USA
;       ZIP:  20001-4598
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/741,940
;       FILING DATE:  19920109
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Kagan, Sarah A.
;       REGISTRATION NUMBER:  32,141
;       REFERENCE/DOCKET NUMBER:  1107.035574
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-508-9100
;       TELEFAX:  202-508-9299
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3172 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
;     IMMEDIATE SOURCE:
;       CLONE:  DP1(TB2)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..630
US-07-741-940-3

  Query Match              1.9%;  Score 33.8;  DB 1;  Length 3172;
  Best Local Similarity   60.2%;  Pred. No. 4.2;
  Matches   56;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy   1022 CTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCAC 1081
          || ||   || ||| |||| |||| |||| |||||| ||| ||| ||    |||
Db    583 CTTCTTTAGTGATGGCATCTGCAGTCTCTTTGGACTTGTCTTTAAGGTCCTTGACCACAC 524

Qy   1082 GGAACATCTACCAGCTTTCCTTCAGACAAAATG 1114
            |  |||||   |    | ||||||  | || |
Db    523 TGTCCATCTGGGACTCGTGCTTCAGGAAGAAAG 491


RESULT 12
US-08-289-548A-3/c
; Sequence 3, Application US/08289548A
; Patent No. 5648212
;  GENERAL INFORMATION:
;    APPLICANT:  ALBERTSEN, HANS
;    APPLICANT:  ANAND, RAKESH
;    APPLICANT:  CARLSON, MARY
;    APPLICANT:  GRODEN, JOANNA
;    APPLICANT:  HEDGE, PHILIP J.
;    APPLICANT:  JOSLYN, GEOFF
;    APPLICANT:  KINZLER, KENNETH
;    APPLICANT:  MARKHAM, ALEXANDER F.
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  THLIVERIS, ANDREW
;    TITLE OF INVENTION:  INHERITED AND SOMATIC MUTATIONS OF APC
;    TITLE OF INVENTION:  GENE IN COLORECTAL CANCER IN HUMANS
;    NUMBER OF SEQUENCES:  102
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner & Allegretti, LTD
;      STREET:  1001 G Street, NW
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20001-4598
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/289,548A
;      FILING DATE:  12-AUG-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kagan, Sarah A.
;      REGISTRATION NUMBER:  32,141
;      REFERENCE/DOCKET NUMBER:  1107.46943
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-508-9100
;      TELEFAX:  202-508-9299
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3172 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  DP1(TB2)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..630
US-08-289-548A-3

  Query Match              1.9%;  Score 33.8;  DB 1;  Length 3172;
  Best Local Similarity   60.2%;  Pred. No. 4.2;
  Matches   56;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy   1022 CTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCAC 1081
          || ||   || ||| |||| |||| |||| |||||| ||| ||| ||    |||
Db    583 CTTCTTTAGTGATGGCATCTGCAGTCTCTTTGGACTTGTCTTTAAGGTCCTTGACCACAC 524

Qy   1082 GGAACATCTACCAGCTTTCCTTCAGACAAAATG 1114
            |  |||||   |    | ||||||  | || |
Db    523 TGTCCATCTGGGACTCGTGCTTCAGGAAGAAAG 491


RESULT 13
US-08-452-654-3/c
; Sequence 3, Application US/08452654
; Patent No. 5691454
;  GENERAL INFORMATION:
;    APPLICANT:  ALBERTSEN, HANS
;    APPLICANT:  ANAND, RAKESH
;    APPLICANT:  CARLSON, MARY
;    APPLICANT:  GRODEN, JOANNA
;    APPLICANT:  HEDGE, PHILIP J.
;    APPLICANT:  JOSLYN, GEOFF
;    APPLICANT:  KINZLER, KENNETH
;    APPLICANT:  MARKHAM, ALEXANDER F.
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  THLIVERIS, ANDREW
;    TITLE OF INVENTION:  INHERITED AND SOMATIC MUTATIONS OF APC
;    TITLE OF INVENTION:  GENE IN COLORECTAL CANCER IN HUMANS
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner, Birch, McKie & Beckett
;      STREET:  1001 G Street, NW
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20001-4598
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,654
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3172 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: DP1(TB2)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..630
US-08-452-654-3

Query Match 1.9%; Score 33.8; DB 1; Length 3172;
Best Local Similarity 60.2%; Pred. No. 4.2;
Matches 56; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 1022 CTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCAC 1081
|| || || ||| |||| |||| |||| |||| ||| ||| || |||
Db 583 CTTCTTTAGTGATGGCATCTGCAGTCTCTTTGGACTTGTCTTTAAGGTCCTTGACCACAC 524

Qy 1082 GGAACATCTACCAGCTTTCCTTCAGACAAAATG 1114
| ||||| | | |||||| | || |
Db 523 TGTCCATCTGGGACTCGTGCTTCAGGAAGAAAG 491

RESULT 14
US-08-452-655B-3/c
; Sequence 3, Application US/08452655B
; Patent No. 5783666
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,655B
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3172 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: DP1(TB2)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..630
US-08-452-655B-3

Query Match 1.9%; Score 33.8; DB 1; Length 3172;
Best Local Similarity 60.2%; Pred. No. 4.2;
Matches 56; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 1022 CTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCAC 1081
|| || || ||| |||| |||| |||| |||| ||| ||| || |||
Db 583 CTTCTTTAGTGATGGCATCTGCAGTCTCTTTGGACTTGTCTTTAAGGTCCTTGACCACAC 524

Qy 1082 GGAACATCTACCAGCTTTCCTTCAGACAAAATG 1114
| ||||| | | |||||| | || |
Db 523 TGTCCATCTGGGACTCGTGCTTCAGGAAGAAAG 491

RESULT 15
US-08-450-582-3/c
; Sequence 3, Application US/08450582
; Patent No. 6114124
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA

; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,582
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/452,655
; FILING DATE: 25-MAY-1995
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3172 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: DP1(TB2)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..630
US-08-450-582-3

Query Match 1.9%; Score 33.8; DB 3; Length 3172;
Best Local Similarity 60.2%; Pred. No. 4.2;
Matches 56; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```
Qy   1022 CTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCAC 1081
          || ||   || ||| |||| |||| ||||    ||||||   ||| ||| ||    |||
Db    583 CTTCTTTAGTGATGGCATCTGCAGTCTCTTTGGACTTGTCTTTAAGGTCCTTGACCACAC 524

Qy   1082 GGAACATCTACCAGCTTTCCTTCAGACAAAATG 1114
          |  |||||   |    | ||||||  | || |
Db    523 TGTCCATCTGGGACTCGTGCTTCAGGAAGAAAG 491
```

Search completed: March 31, 2003, 18:30:43
Job time : 146 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 31, 2003, 15:07:07 ; Search time 396 Seconds
(without alignments)
9912.205 Million cell updates/sec

Title: US-10-069-541-5
Perfect score: 1743
Sequence: 1 atggctttccatgtggaagg.....ctgaagataatttacagtga 1743

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1743	100.0	1743	22	AAF81712	Human high affinit
2	1743	100.0	1743	22	AAH49207	Human CHOT encodin
3	1394.2	80.0	1743	22	AAF81711	Rat high affinity
4	1373.4	78.8	1743	22	AAF81713	Mouse high affinit
5	1373.4	78.8	4938	22	AAD02457	Mouse P4P6B1 OMA (
6	630.8	36.2	8760	22	AAH49206	Human CHOT exons 6
7	363.8	20.9	1731	22	AAF81710	C. elegans high af
8	279.6	16.0	386	22	AAD02461	Mouse P4P6B1 cDNA
9	266.6	15.3	1729	23	ABL29569	Drosophila melanog
10	242.6	13.9	1461	24	ABK73210	Bacillus lichenifo
11	180.8	10.4	10140	22	AAH49201	Human CHOT promote
c 12	179.6	10.3	4223	23	ABL29568	Drosophila melanog
13	167.6	9.6	1094	23	AAS82193	DNA encoding novel
14	163	9.4	240	22	AAH49203	Human CHOT exon 3
15	155	8.9	455	22	ABA53620	Human foetal liver
16	155	8.9	455	22	ABA23372	Probe #1838 for ge
17	155	8.9	455	22	AAK01883	Human brain expres
18	155	8.9	455	22	AAK27341	Human bone marrow
19	155	8.9	455	22	AAI11913	Probe #1846 for ge
20	155	8.9	455	22	AAI33245	Probe #1931 used t
21	155	8.9	455	22	AAI01851	Probe #1842 used t
22	154.6	8.9	300	22	AAH49204	Human CHOT exon 4
23	150.8	8.7	240	22	AAH49205	Human CHOT exon 5
24	108	6.2	180	22	AAH49202	Human CHOT exon 2
25	72	4.1	96	22	ABA66205	Human foetal liver
26	72	4.1	96	22	ABA33269	Probe #11735 for g
27	72	4.1	96	22	AAK14621	Human brain expres
28	72	4.1	96	22	AAK40366	Human bone marrow
29	72	4.1	96	22	AAI21121	Probe #11054 for g
30	72	4.1	96	22	AAI46388	Probe #15074 used
31	72	4.1	96	22	AAI06841	Probe #6832 used t
32	60	3.4	60	24	ABN37501	Human spliced tran
33	53.8	3.1	65	24	ABN54094	Mouse spliced tran
c 34	51.8	3.0	10732	21	AAA10594	Gene encoding a su
35	42.4	2.4	349980	22	AAF86431	Pyrococcus abyssi
c 36	41.6	2.4	11220	21	AAC48974	Arabidopsis thalia
37	41.2	2.4	2028	24	ABL57735	Human sbg1007026SG
38	41.2	2.4	2200	22	AAF27713	Human transport pr
39	41	2.4	1857	22	AAF79537	Rat NIS nucleotide
40	41	2.4	2839	18	AAT88971	Rat sodium/iodide
c 41	39.8	2.3	4590	22	AAH24065	Yeast AOD9604-asso
42	39.6	2.3	2028	22	AAF84022	Novel human transp
43	39.6	2.3	2456	22	AAF84023	Novel human transp
44	38.2	2.2	2326	24	ABK10764	Human sodium-sugar
45	38.2	2.2	6426	21	AAA57920	Candida albicans p

ALIGNMENTS

RESULT 1
AAF81712
ID AAF81712 standard; cDNA; 1743 BP.
XX
AC AAF81712;
XX
DT 01-JUN-2001 (first entry)
XX
DE Human high affinity choline transporter protein encoding cDNA.
XX
KW High affinity choline transporter; cho-1; Alzheimer's disease;
KW diagnosis; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1743
FT /*tag= a
FT /product= "high affinity choline transporter"
FT
XX
PN WO200116315-A1.
XX
PD 08-MAR-2001.
XX
PF 18-AUG-2000; 2000WO-JP05545.
XX
PR 27-AUG-1999; 99JP-0240642.
PR 27-DEC-1999; 99JP-0368991.
XX
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI Haga T, Okuda T;

XX
DR WPI; 2001-226688/23.
DR P-PSDB; AAB74665.
XX
PT New rat and human spinal cord high affinity choline transporters,
PT useful in diagnosis of Alzheimer's disease and screening promoters as
PT drugs for treating Alzheimer's disease -
XX
PS Claim 9; Page 71-75; 90pp; Japanese.
XX
CC The present sequence encodes a human (Homo sapiens) high affinity
CC choline transporter protein designated cho-1. The cho-1 protein has
CC nootropic and neuroprotective activities. The cho-1 polynucleotide and
CC protein can be used for the diagnosis of diseases related to the
CC expression of cho-1 by comparing the cho-1 polynucleotide sequence in a
CC sample to that of a control. Drug compositions containing the cho-1
CC protein or expression promoters or inhibitors of cho-1 are useful for
CC treating disorders characterised by abnormal levels of cho-1, such as
CC Alzheimer's disease.
XX
SQ Sequence 1743 BP; 412 A; 393 C; 406 G; 532 T; 0 other;

Query Match 100.0%; Score 1743; DB 22; Length 1743;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1743; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60

Qy     61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120

Qy    121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180

Qy    181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240

Qy    241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300

Qy    301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360

Qy    361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420

Qy    421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 480

Qy    481 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 540

Qy    541 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 600

Qy    601 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 660

Qy    661 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 720

Qy    721 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 780

Qy    781 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 840

Qy    841 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 900

Qy    901 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 960

Qy    961 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1020

Qy   1021 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1080

Qy   1081 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140

Qy   1141 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1200

Qy   1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1260

Qy   1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1320

Qy   1321 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1380

Qy   1381 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1440

Qy   1441 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1500

Qy   1501 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1560

Qy   1561 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1620

Qy   1621 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1680

Qy   1681 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1740

Qy   1741 TGA 1743
          |||
Db   1741 TGA 1743
```

RESULT 2
AAH49207
ID AAH49207 standard; cDNA; 1743 BP.
XX
AC AAH49207;
XX
DT 26-NOV-2001 (first entry)
XX
DE Human CHOT encoding cDNA.
XX
KW CHOT; human; choline transporter; chromosome 2q11-13; nootropic;
KW neuroprotective; gene therapy; antisense therapy; degenerative disease;
KW cognitive disorder; Alzheimer's disease; ss.
XX
OS Homo sapiens.
XX
PN DE10009055-A1.
XX
PD 30-AUG-2001.
XX
PF 28-FEB-2000; 2000DE-1009055.
XX
PR 28-FEB-2000; 2000DE-1009055.
XX
PA (BRUE/) BRUESS M.
PA (BOEN/) BOENISCH H.
XX
PI Bruess M, Boenisch H;
XX
DR WPI; 2001-590709/67.
DR P-PSDB; AAB86837.
XX
PT A new gene encoding human choline transporter, designated hCHOT is
PT located on chromosome 2q11-13 and is useful to treat degenerative
PT disorders such as Alzheimer's disease
XX
PS Disclosure; Page 11; 12pp; German.
XX
CC This invention describes a novel gene encoding human choline transporter,
CC designated hCHOT which is located on chromosome 2q11-13. The products of
CC the invention have nootropic and neuroprotective activity and can be used
CC for gene or antisense therapy. (I) is used to treat degenerative disease,
CC particularly cognitive disorders such as Alzheimer's disease. Sense and
CC antisense oligonucleotides derived from the gene may be used in
CC diagnostics and other techniques. This sequence encodes the human CHOT
CC protein described in the invention.
XX
SQ Sequence 1743 BP; 412 A; 393 C; 406 G; 532 T; 0 other;

Query Match 100.0%; Score 1743; DB 22; Length 1743;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1743; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60
Db       1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60

Qy      61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120
Db      61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120

Qy     121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180
Db     121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180

Qy     181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240
Db     181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240

Qy     241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300
Db     241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300

Qy     301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360
Db     301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360

Qy     361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420
Db     361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420

Qy     421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 480
Db     421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 480

Qy     481 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 540
Db     481 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 540

Qy     541 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 600
Db     541 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 600

Qy     601 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 660
Db     601 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 660

Qy     661 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 720
Db     661 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 720

Qy     721 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 780
Db     721 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 780

Qy     781 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 840
Db     781 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 840

Qy     841 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 900
Db     841 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 900

Qy     901 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 960
Db     901 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 960

Qy     961 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1020
Db     961 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1020

Qy    1021 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1080
Db    1021 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1080

Qy    1081 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140
Db    1081 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140

Qy    1141 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1200
Db    1141 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1200

Qy    1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1260
Db    1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1260

Qy    1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1320
Db    1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1320

Qy    1321 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1380
Db    1321 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1380
```

```
Qy    1381 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1440

Qy    1441 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1500

Qy    1501 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1560

Qy    1561 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1620

Qy    1621 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1680

Qy    1681 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1740

Qy    1741 TGA 1743
           |||
Db    1741 TGA 1743
```

RESULT 3
AAF81711

```
ID   AAF81711 standard; cDNA; 1743 BP.
XX
AC   AAF81711;
XX
DT   01-JUN-2001  (first entry)
XX
DE   Rat high affinity choline transporter protein encoding cDNA.
XX
KW   High affinity choline transporter; cho-1; Alzheimer's disease;
KW   diagnosis; ss.
XX
OS   Rattus norvegicus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1743
FT                   /*tag=  a
FT                   /product= "high affinity choline transporter"
XX
PN   WO200116315-A1.
XX
PD   08-MAR-2001.
XX
PF   18-AUG-2000; 2000WO-JP05545.
XX
PR   27-AUG-1999;   99JP-0240642.
PR   27-DEC-1999;   99JP-0368991.
XX
PA   (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI   Haga T,  Okuda T;
XX
DR   WPI; 2001-226688/23.
DR   P-PSDB; AAB74664.
XX
PT   New rat and human spinal cord high affinity choline transporters,
PT   useful in diagnosis of Alzheimer's disease and screening promoters as
PT   drugs for treating Alzheimer's disease  -
XX
PS   Claim 6; Page 64-68; 90pp; Japanese.
XX
CC   The present sequence encodes a rat (Rattus norvegicus) high affinity
CC   choline transporter protein designated cho-1. The cho-1 protein has
CC   nootropic and neuroprotective activities. The cho-1 polynucleotide and
CC   protein can be used for the diagnosis of diseases related to the
CC   expression of cho-1 by comparing the cho-1 polynucleotide sequence in a
CC   sample to that of a control. Drug compositions containing the cho-1
CC   protein or expression promoters or inhibitors of cho-1 are useful for
CC   treating disorders characterised by abnormal levels of cho-1, such as
CC   Alzheimer's disease.
XX
SQ   Sequence 1743 BP; 414 A; 402 C; 404 G; 523 T; 0 other;
```

Query Match 80.0%; Score 1394.2; DB 22; Length 1743;
Best Local Similarity 87.5%; Pred. No. 0;
Matches 1525; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

```
Qy       1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60
           ||| |||||||||| ||||||||  |||| || ||| ||||||||||||| || || |||
Db       1 ATGCCTTTCCATGTAGAAGGACTAGTAGCGATTATCCTGTTCTACCTTCTTATATTTCTG 60

Qy      61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120
           ||||||||||||||||| |||| |||||||||||| || |  |||||||| |||||||||
Db      61 GTTGGAATATGGGCTGCATGGAAAACCAAAAACAGCGGTAATGCAGAAGAACGCAGCGAA 120

Qy     121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180
           ||||||||||||| |||||||| |||||||| |||||||||| ||||||||||||||||
Db     121 GCCATCATAGTTGGGGGCCGAGACATTGGTTTGTTGGTTGGTGGTTTTACCATGACAGCC 180

Qy     181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240
           |||||||| |||||||| || ||||| || ||||||||||||||||||  ||||||| |
Db     181 ACCTGGGTTGGAGGAGGTTACATCAACGGGACAGCTGAAGCAGTTTATGGGCCAGGTTGT 240

Qy     241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300
           || |||||||||||||||||||| |||||||||||||| ||||||||||||||||||||||
Db     241 GGTCTAGCTTGGGCTCAGGCACCCATTGGATATTCTCTGAGTCTGATTTTAGGTGGCCTG 300

Qy     301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360
           || |||||||||||||||||||| ||||| |||||||| ||||||||||||||||||| ||
Db     301 TTTTTTGCAAAACCTATGCGTTCCAAGGGATATGTGACTATGTTAGACCCGTTTCAACAG 360

Qy     361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420
           |||||||||||| |||||||| || || ||||| || |||||||||||||||||| ||||||
Db     361 ATCTATGGAAAGCGCATGGGTGGGCTGCTGTTCATCCCTGCACTGATGGGAGAGATGTTC 420

Qy     421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 480
           |||||||||||||||||||||||| || || || |||||||||| ||||| |||||||||
Db     421 TGGGCTGCAGCAATTTTCTCTGCATTAGGGGCTACCATCAGCGTAATCATTGATGTGGAT 480

Qy     481 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 540
           || |||| || |||||  |||| |||||||||||||| ||| || || || |||||||||
Db     481 GTGAACATATCGGTCATTGTCTCCGCACTCATTGCCATTCTTTATACCCTCGTGGGAGGG 540

Qy     541 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 600
           ||||| |||||||| || |||||||| || ||||| || |||||||||  ||||  |||||
Db     541 CTCTACTCTGTGGCATATACTGATGTTGTACAGCTATTCTGCATTTTTATAGGATTGTGG 600

Qy     601 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 660
           ||||| ||||| |||||  |||||||||||||||| | ||||| || ||||||||||||
Db     601 ATCAGTGTCCCATTTGCCCTGTCACATCCTGCAGTCACCGACATTGGATTCACTGCTGTG 660

Qy     661 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 720
           ||||| |||||||| |  || ||||||||||  |||| |||  |||||||||| | |||
Db     661 CATGCTAAATACCAGAGTCCCTGGCTGGGAACCATTGAATCAGTTGAAGTCTACACCTGG 720

Qy     721 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 780
           ||||||| |||||||||||||||||||||||||||||||||| ||||||||||| |||||
Db     721 CTTGATAATTTTCTGTTGTTGATGCTGGGTGGAATACCATGGCAAGCCTACTTCCAGAGG 780

Qy     781 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 840
           || |||||||| || ||||| |||||||||| |||||||||||||||||||||||| |||
Db     781 GTCCTCTCTTCATCGTCAGCGACCTATGCTCAGGTGCTGTCCTTCCTGGCAGCTTTTGGG 840

Qy     841 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 900
           ||||||||||||||  | ||||||||  |||||||||||||||||| || |||||||||
Db     841 TGCCTGGTGATGGCTCTACCAGCCATTTGCATTGGGGCCATTGGAGCCTCCACAGACTGG 900
```

```
Qy    901 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 960
          ||||| |||||||||||| |||||||||||||||| ||  || || ||||||||||||
Db    901 AACCAAACTGCATATGGGTTTCCAGATCCCAAGACCAAGGAGGAAGCAGACATGATTCTC 960

Qy    961 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1020
          || |||||||| ||||| |||||||||||||| ||||| |||||||| |||||||| |||
Db    961 CCGATTGTTCTACAGTACCTCTGCCCTGTGTACATTTCCTTCTTTGGGCTTGGTGCTGTT 1020

Qy   1021 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1080
          |||||||||| ||||| || || || || |||||| | ||||||||||||||||||||
Db   1021 TCTGCTGCTGTCATGTCCTCGGCTGACTCATCCATCCTATCAGCAAGTTCCATGTTTGCT 1080

Qy   1081 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140
          ||||| |||||||||||||||||||||||||| || ||||| ||||| || |||||
Db   1081 CGGAATATCTACCAGCTTTCCTTCAGACAAAATGCATCAGACAAGGAAATTGTGTGGGTC 1140

Qy   1141 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1200
          ||| | ||||| |||||||||||||||||||||||||||||||||||||||||| |||||
Db   1141 ATGAGGATCACTGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTCACGAAG 1200

Qy   1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1260
          ||||||||||||||||||||||| || |||||||||| |||||| | |||||||| |||
Db   1201 ACTGTGTATGGGCTCTGGTACCTGAGCTCTGACCTTGTCTACATCATCATCTTCCCACAG 1260

Qy   1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1320
          ||||| ||||||||||  | || |||||||||| |||||||| || || |||||| ||
Db   1261 CTGCTCTGTGTACTCTTCATCAAAGGAACCAACACTTATGGGGCAGTTGCTGGTTATATT 1320

Qy   1321 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1380
          | ||| || |||||||||| || ||||| |||||||||| ||  | |||||||| |||
Db   1321 TTTGGACTTTTCCTGAGAATTACCGGAGGAGAGCCATATCTATACTTGCAGCCCTTAATC 1380

Qy   1381 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1440
          |||||||||| ||||||||||  | |||||||||| ||||||| || |||||||||
Db   1381 TTCTACCCTGGTTATTACCCTGACAAGAATGGTATATACAATCAGAGGTTCCCATTTAAA 1440

Qy   1441 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1500
          || ||  |||||||||| |||||||| |||||||||||  | ||||||||||||||||||
Db   1441 ACTCTCTCCATGGTTACCTCATTCTTTACCAACATTTGTGTTTCCTATCTAGCCAAGTAT 1500

Qy   1501 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1560
          |||||||||||||||||||||||| || ||||||||| |||||||||||||||  ||||
Db   1501 CTATTTGAAAGTGGAACCTTGCCTCCAAAATTAGATATATTTGATGCTGTTGTCTCAAGG 1560

Qy   1561 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1620
          |||||||||| |||||||| ||||| ||||| |||| |||||||| || |||||| ||
Db   1561 CACAGTGAAGAGAACATGGACAAGACCATTCTAGTCAGAAATGAAAACATCAAATTAAAT 1620

Qy   1621 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1680
          |||||||||| ||| ||||| |||||||| | |||||||| ||||||||||||||||||
Db   1621 GAACTTGCACCTGTAAAGCCTCGACAGAGCCTAACCCTCAGTTCAACTTTCACCAATAAA 1680

Qy   1681 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1740
          |||||  |||||||||||||||||||||||| || ||||||||||||||||  |||||
Db   1681 GAGGCTCTCCTTGATGTTGATTCCAGTCCAGAGGGATCTGGGACTGAAGATAACTTACAA 1740

Qy   1741 TGA 1743
          |||
Db   1741 TGA 1743
```

```
RESULT 4
AAF81713
ID   AAF81713 standard; cDNA; 1743 BP.
XX
AC   AAF81713;
XX
DT   01-JUN-2001  (first entry)
XX
DE   Mouse high affinity choline transporter protein encoding cDNA.
XX
KW   High affinity choline transporter; cho-1; Alzheimer's disease;
KW   diagnosis; ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1743
FT                   /*tag= a
FT                   /product= "high affinity choline transporter"
XX
PN   WO200116315-A1.
XX
PD   08-MAR-2001.
XX
PF   18-AUG-2000; 2000WO-JP05545.
XX
PR   27-AUG-1999;   99JP-0240642.
PR   27-DEC-1999;   99JP-0368991.
XX
PA   (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI   Haga T,  Okuda T;
XX
DR   WPI; 2001-226688/23.
DR   P-PSDB; AAB74666.
XX
PT   New rat and human spinal cord high affinity choline transporters,
PT   useful in diagnosis of Alzheimer's disease and screening promoters as
PT   drugs for treating Alzheimer's disease -
XX
PS   Claim 12; Page 78-82; 90pp; Japanese.
XX
CC   The present sequence encodes a mouse (Mus musculus) high affinity
CC   choline transporter protein designated cho-1. The cho-1 protein has
CC   nootropic and neuroprotective activities. The cho-1 polynucleotide and
CC   protein can be used for the diagnosis of diseases related to the
CC   expression of cho-1 by comparing the cho-1 polynucleotide sequence in a
CC   sample to that of a control. Drug compositions containing the cho-1
CC   protein or expression promoters or inhibitors of cho-1 are useful for
CC   treating disorders characterised by abnormal levels of cho-1, such as
CC   Alzheimer's disease.
XX
SQ   Sequence 1743 BP; 407 A; 410 C; 409 G; 517 T; 0 other;

  Query Match             78.8%;  Score 1373.4;  DB 22;  Length 1743;
  Best Local Similarity   86.7%;  Pred. No. 0;
  Matches 1512;  Conservative    0;  Mismatches  231;  Indels    0;  Gaps    0;

Qy      1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60
          ||| ||||||| || ||||||||| ||||||| ||| | |||||||| || || || |||
Db      1 ATGTCTTTCCACGTAGAAGGACTGGTAGCTATTATCCTCTTCTACCTCCTTATATTTCTG 60

Qy     61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120
          |||||||||||||||| |||| |||||||||||| |||| | |||||||| ||| |||
Db     61 GTTGGAATATGGGCTGCATGGAAAACCAAAAACAGCGGCAACCCAGAAGAGCACAGTGAA 120

Qy    121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180
          |||||||||| || ||||| || |||||||| |||||||||||  ||||||||||||||
Db    121 GCCATCATAGTCGGGGGCCGTGACATTGGTTTGTTGGTTGGTGGTTTTACCATGACAGCC 180

Qy    181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240
          |||||||| |||||||| || |||||||| ||||| |||||||| ||||  ||||||| |
Db    181 ACCTGGGTTGGAGGAGGCTACATCAATGGGACAGCAGAAGCAGTGTATGGGCCAGGTTGT 240

Qy    241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300
          || |||||||||||||||||| |||||||||||||| ||||| |||||||||||  |||
Db    241 GGTCTAGCTTGGGCTCAGGCACCCATTGGATATTCTCTGAGTCTAATTTTAGGTGGTCTG 300

Qy    301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360
          || ||||| |||||||||||||| ||||| |||||||| |||||||||||  ||||| ||
Db    301 TTTTTTGCGAAACCTATGCGTTCCAAGGGATATGTGACTATGTTAGACCCATTTCAACAG 360

Qy    361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420
          |||||||||||  |||||||| || || || || || ||||||||||||||||||  ||||||
```

```
Db    361 ATCTATGGAAAGCGCATGGGTGGGCTGCTCTTCATCCCTGCACTGATGGGAGAGATGTTC 420
Qy    421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 480
Db    421 TGGGCTGCAGCAATTTTCTCTGCATTAGGGGCCACCATCAGCGTGATCATTGATGTGGAT 480
Qy    481 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 540
Db    481 GTGAACATATCGGTCATTGTCTCTGCACTCATTGCCATTCTTTATACCCTAGTGGGTGGG 540
Qy    541 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 600
Db    541 CTCTACTCTGTGGCATATACTGATGTTGTCCAGCTATTCTGCATTTTTATAGGACTGTGG 600
Qy    601 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 660
Db    601 ATCAGTGTCCCTTTTGCCCTGTCACATCCTGCAGTCACCGACATCGGATTCACAGCTGTG 660
Qy    661 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 720
Db    661 CATGCTAAATACCAGAGTCCCTGGCTGGGAACCATTGAATCAGTTGAAGTCTACACCTGG 720
Qy    721 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 780
Db    721 CTTGATAATTTTCTGTTATTGATGCTGGGTGGAATCCCATGGCAAGCCTACTTCCAGAGG 780
Qy    781 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 840
Db    781 GTCCTCTCTTCATCCTCAGCCACCTATGCTCAGGTACTGTCCTTCCTGGCAGCTTTTGGG 840
Qy    841 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 900
Db    841 TGCCTGGTGATGGCTCTACCCGCCATATGCATAGGAGCTATTGGAGCTTCCACAGACTGG 900
Qy    901 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 960
Db    901 AACCAGACTGCCTACGGGTATCCAGATCCCAAGACTAAGGAGGAAGCAGACATGATTCTC 960
Qy    961 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1020
Db    961 CCGATCGTTCTGCAGTACCTCTGCCCTGTGTACATCTCCTTCTTTGGGCTTGGTGCTGTT 1020
Qy   1021 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1080
Db   1021 TCAGCTGCTGTCATGTCCTCAGCTGACTCGTCCATCCTGTCGGCGAGTTCTATGTTTGCT 1080
Qy   1081 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140
Db   1081 CGGAATATCTACCAGCTTTCCTTCAGACAAAATGCATCAGACAAGGAAATTGTGTGGGTC 1140
Qy   1141 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1200
Db   1141 ATGAGGATCACTGTGCTTGTGTTCGGAGCATCTGCAACAGCCATGGCTTTGCTGACGAAG 1200
Qy   1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1260
Db   1201 ACTGTGTATGGGCTCTGGTACCTGAGCTCTGACCTTGTCTACATCATCATCTTCCCACAG 1260
Qy   1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1320
Db   1261 CTGCTCTGTGTACTCTTCATCAAAGGAACCAACACTTATGGGGCAGTTGCTGGTTATATT 1320
Qy   1321 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1380
Db   1321 TTTGGACTATTCCTGAGAATTACTGGAGGAGAGCCATATCTATACTTGCAGCCCTTAATC 1380
Qy   1381 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1440
Db   1381 TTCTACCCTGGTTATTACTCTGACAAGAATGGTATATACAATCAGAGGTTCCCATTTAAA 1440
Qy   1441 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1500
Db   1441 ACTCTCTCCATGGTTACCTCATTCTTTACCAACATTTGTGTTTCTTATCTAGCCAAGTAT 1500
Qy   1501 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1560
Db   1501 CTATTTGAAAGTGGAACCTTGCCTCCAAAATTAGATGTATTTGATGCTGTTGTCGCAAGG 1560
Qy   1561 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1620
Db   1561 CACAGTGAAGAGAACATGGACAAGACCATTCTAGTCAGAAATGAAAATATCAAATTAAAT 1620
Qy   1621 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1680
Db   1621 GAACTTGCACCTGTGAAACCTCGGCAGAGCCTAACCCTCAGTTCAACTTTCACCAATAAG 1680
Qy   1681 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1740
Db   1681 GAGGCCCTCCTTGATGTTGATTCCAGTCCGGAGGGGTCTGGGACTGAAGATAACTTACAA 1740
Qy   1741 TGA 1743
Db   1741 TGA 1743
```

RESULT 5
AAD02457
ID AAD02457 standard; cDNA; 4938 BP.
XX
AC AAD02457;
XX
DT 24-APR-2001 (first entry)
XX
DE Mouse P4P6B1 OMA (obese mice adipocyte) protein encoding cDNA.
XX
KW Mouse; OMA protein; obese mice adipocyte; P4P6B1;
KW fuel metabolism disorder; therapy; obesity; diabetes; gene therapy;
KW anorectic; antidiabetic; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 247..1989
FT /*tag= a
FT /product= "Mouse P4P6B1 OMA protein"
FT misc_feature 988..1342
FT /*tag= b
FT /note= "Portion of original 450 bp PCR fragment"
XX
PN WO200078950-A2.
XX
PD 28-DEC-2000.
XX
PF 13-JUN-2000; 2000WO-US16217.
XX
PR 22-JUN-1999; 99US-0141515.
XX
PA (AMYL-) AMYLIN PHARM INC.
XX
PI Sierzega M, Albrandt K;
XX
DR WPI; 2001-112322/12.
DR P-PSDB; AAY72388.
XX
PT Novel obese mice adipocyte polypeptides useful in diagnosis and
PT treatment of disorders of fuel metabolism such as obesity or diabetes
PT -
XX
PS Claim 2; Fig 3; 83pp; English.
XX
CC The present sequence is mouse P4P6B1 cDNA which encodes OMA (obese
CC mice adipocyte) protein. The P4P6B1 fragment was generated by RNA
CC fingerprinting using random primers P4 and P6.
CC OMA is used as a diagnostic reagent for diagnosing a disorder of fuel
CC metabolism in an underweight or an overweight individual, by detecting
CC the transcription level of a gene encoding OMA, which is induced or

CC repressed in an individual by a factor such as genetic obesity, fasting
CC and refeeding of a fasted individual. OMA is useful in the generation of
CC antibodies, for use in pharmaceutical compositions and for studying
CC DNA/protein interactions. Nucleic acids encoding OMA are involved in
CC gene therapy. An inhibitor of OMA or an antisense oligonucleotide that
CC inhibits expression of OMA are useful for treating disorders of fuel
CC metabolism such as obesity or diabetes.
XX
SQ Sequence 4938 BP; 1436 A; 1012 C; 976 G; 1514 T; 0 other;

Query Match 78.8%; Score 1373.4; DB 22; Length 4938;
Best Local Similarity 86.7%; Pred. No. 0;
Matches 1512; Conservative 0; Mismatches 231; Indels 0; Gaps 0;

```
Qy      1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60
          ||| ||||||| || |||||||| ||||||| ||| | |||||||| || ||  | |||
Db    247 ATGTCTTTCCACGTAGAAGGACTGGTAGCTATTATCCTCTTCTACCTCCTTATACTTCTG 306

Qy     61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120
          |||||||||||||||| |||| |||||||||||| |||| | ||||||||||||| |||
Db    307 GTTGGAATATGGGCTGCATGGAAAACCAAAAACAGCGGCAACCCAGAAGAGCGCAGTGAA 366

Qy    121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180
          |||||||||| || |||| || |||||||| |||||||||| ||||||||||||
Db    367 GCCATCATAGTCGGGGGCCGTGACATTGGTTTGTTGGTTGGTGGTTTTACCATGACAGCC 426

Qy    181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240
          |||||||| |||||||| || |||||||| |||| |||||||| ||||  ||||||| |
Db    427 ACCTGGGTTGGAGGAGGCTACATCAATGGGACAGCAGAAGCAGTGTATGGGCCAGGTTGT 486

Qy    241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300
          || |||||||||||||||||||| |||||||||||||| ||||| |||||||||||| |||
Db    487 GGTCTAGCTTGGGCTCAGGCACCCATTGGATATTCTCTGAGTCTAATTTTAGGTGGTCTG 546

Qy    301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360
          || ||||| |||||||||||||| ||||| |||||||| |||||||||||| ||||| ||
Db    547 TTTTTTGCGAAACCTATGCGTTCCAAGGGATATGTGACTATGTTAGACCCATTTCAACAG 606

Qy    361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420
          |||||||||| |||||||| || || || || || |||||||||||||||||| ||||||
Db    607 ATCTATGGAAAGCGCATGGGTGGGCTGCTCTTCATCCCTGCACTGATGGGAGAGATGTTC 666

Qy    421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 480
          |||||||||||||||||||||||||| || || |||||||||||||||||||| |||||||||
Db    667 TGGGCTGCAGCAATTTTCTCTGCATTAGGGGCCACCATCAGCGTGATCATTGATGTGGAT 726

Qy    481 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 540
          || |||| || |||||  |||||||||||||||||| ||| || || || ||||| |||
Db    727 GTGAACATATCGGTCATTGTCTCTGCACTCATTGCCATTCTTTATACCCTAGTGGGTGGG 786

Qy    541 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 600
          ||||| |||||||| || |||||||| || ||||| || |||||||| |||| ||||||
Db    787 CTCTACTCTGTGGCATATACTGATGTTGTCCAGCTATTCTGCATTTTTATAGGACTGTGG 846

Qy    601 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 660
          ||||| ||||| |||||  |||||||||||||||||| | |||||||| ||||| ||||||
Db    847 ATCAGTGTCCCTTTTGCCCTGTCACATCCTGCAGTCACCGACATCGGATTCACAGCTGTG 906

Qy    661 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 720
          ||||| |||||||| |  || |||||||||| |||| |||  |||||||||| | |||
Db    907 CATGCTAAATACCAGAGTCCCTGGCTGGGAACCATTGAATCAGTTGAAGTCTACACCTGG 966

Qy    721 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 780
          ||||||| ||||||||| ||||||||||||||||||||||||||||||| ||||| ||||||
Db    967 CTTGATAATTTTCTGTTATTGATGCTGGGTGGAATCCCATGGCAAGCCTACTTCCAGAGG 1026

Qy    781 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 840
          || |||||||| |||||||||||||||||||||| || ||||||||||||||||||| |||
Db   1027 GTCCTCTCTTCATCCTCAGCCACCTATGCTCAGGTACTGTCCTTCCTGGCAGCTTTTGGG 1086

Qy    841 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 900
          |||||||||||||  | || ||||||  ||| || || |||||||| || |||||||||
Db   1087 TGCCTGGTGATGGCTCTACCCGCCATATGCATAGGAGCTATTGGAGCTTCCACAGACTGG 1146

Qy    901 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 960
          |||||||||| || |||  |||||||||||||||||  || || |||||||||||| |
Db   1147 AACCAGACTGCCTACGGGTATCCAGATCCCAAGACTAAGGAGGAAGCAGACATGATTCTC 1206

Qy    961 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1020
          || || |||||||||| |||||||||||||| || || |||||||| ||||||| |||
Db   1207 CCGATCGTTCTGCAGTACCTCTGCCCTGTGTACATCTCCTTCTTTGGGCTTGGTGCTGTT 1266

Qy   1021 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1080
          || |||||||| |||| |||| || || |||||| |||| || ||||| ||||||||
Db   1267 TCAGCTGCTGTCATGTCCTCAGCTGACTCGTCCATCCTGTCGGCGAGTTCTATGTTTGCT 1326

Qy   1081 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140
          ||||| |||||||||||||||||||||||||||| || ||||| ||||| || ||||||
Db   1327 CGGAATATCTACCAGCTTTCCTTCAGACAAAATGCATCAGACAAGGAAATTGTGTGGGTC 1386

Qy   1141 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1200
          ||| | ||||| ||| ||||||| |||||||||||||||||||||||||| ||||||||||
Db   1387 ATGAGGATCACTGTGCTTGTGTTCGGAGCATCTGCAACAGCCATGGCTTTGCTGACGAAG 1446

Qy   1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1260
          |||||||||||||||||||||||| || |||||||||| |||||| | |||||||| |||
Db   1447 ACTGTGTATGGGCTCTGGTACCTGAGCTCTGACCTTGTCTACATCATCATCTTCCCACAG 1506

Qy   1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1320
          ||||| |||||||||||  | || |||||||||| |||||||| || || |||||| ||
Db   1507 CTGCTCTGTGTACTCTTCATCAAAGGAACCAACACTTATGGGGCAGTTGCTGGTTATATT 1566

Qy   1321 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1380
          | ||| || |||||||||| |||||||| |||||||||| ||  | |||||||| |||
Db   1567 TTTGGACTATTCCTGAGAATTACTGGAGGAGAGCCATATCTATACTTGCAGCCCTTAATC 1626

Qy   1381 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1440
          |||||||||| |||||| |||| | |||||||||| |||||||  || |||||||||
Db   1627 TTCTACCCTGGTTATTACTCTGACAAGAATGGTATATACAATCAGAGGTTCCCATTTAAA 1686

Qy   1441 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1500
          || ||  |||||||||| |||||||| ||||||||||  | || ||||||||||||||||
Db   1687 ACTCTCTCCATGGTTACCTCATTCTTTACCAACATTTGTGTTTCTTATCTAGCCAAGTAT 1746

Qy   1501 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1560
          |||||||||||||||||||||| || |||||||||||||||||||||||||||| |||||
Db   1747 CTATTTGAAAGTGGAACCTTGCCTCCAAAATTAGATGTATTTGATGCTGTTGTCGCAAGG 1806

Qy   1561 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1620
          |||||||||||  |||||||| ||||| ||||| |||| |||||||||||| |||||| ||
Db   1807 CACAGTGAAGAGAACATGGACAAGACCATTCTAGTCAGAAATGAAAATATCAAATTAAAT 1866

Qy   1621 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1680
          |||||||||| |||||| || || |||||||| |||||||||||||||||||||||||||
Db   1867 GAACTTGCACCTGTGAAACCTCGGCAGAGCCTAACCCTCAGTTCAACTTTCACCAATAAG 1926

Qy   1681 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1740
          |||||| |||||||||||||||||||||| || |||||||||||||||||||| |||||
Db   1927 GAGGCCCTCCTTGATGTTGATTCCAGTCCGGAGGGGTCTGGGACTGAAGATAACTTACAA 1986

Qy   1741 TGA 1743
          |||
Db   1987 TGA 1989
```

RESULT 6
AAH49206
ID AAH49206 standard; DNA; 8760 BP.
XX
AC AAH49206;
XX
DT 26-NOV-2001 (first entry)
XX
DE Human CHOT exons 6, 7, 8 and 3' UTR region DNA.
XX
KW CHOT; human; choline transporter; chromosome 2q11-13; nootropic;

```
KW   neuroprotective; gene therapy; antisense therapy; degenerative disease;
KW   cognitive disorder; Alzheimer's disease; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..4853
FT                   /*tag= a
FT                   /product= "CHOT"
FT                   /note= "This sequence is interrupted by introns"
FT   exon            41..194
FT                   /*tag= b
FT                   /number= "6"
FT   intron          195..2456
FT                   /*tag= c
FT                   /number= "6"
FT   exon            2455..2674
FT                   /*tag= d
FT                   /number= "7"
FT   intron          2675..4223
FT                   /*tag= e
FT                   /number= "7"
FT   exon            4224..4853
FT                   /*tag= f
FT                   /number= "8"
FT   3'UTR           4854..8760
FT                   /*tag= g
XX
PN   DE10009055-A1.
XX
PD   30-AUG-2001.
XX
PF   28-FEB-2000; 2000DE-1009055.
XX
PR   28-FEB-2000; 2000DE-1009055.
XX
PA   (BRUE/) BRUESS M.
PA   (BOEN/) BOENISCH H.
XX
PI   Bruess M,  Boenisch H;
XX
DR   WPI; 2001-590709/67.
XX
PT   A new gene encoding human choline transporter, designated hCHOT is
PT   located on chromosome 2q11-13 and is useful to treat degenerative
PT   disorders such as Alzheimer's disease -
XX
PS   Disclosure; Page 9-11; 12pp; German.
XX
CC   This invention describes a novel gene encoding human choline transporter,
CC   designated hCHOT which is located on chromosome 2q11-13. The products of
CC   the invention have nootropic and neuroprotective activity and can be used
CC   for gene or antisense therapy. (I) is used to treat degenerative disease,
CC   particularly cognitive disorders such as Alzheimer's disease. Sense and
CC   antisense oligonucleotides derived from the gene may be used in
CC   diagnostics and other techniques. This sequence represents exons 6-8 and
CC   the 3' UTR fragment encoding the human CHOT protein described in the
CC   invention.
XX
SQ   Sequence 8760 BP; 2727 A; 1619 C; 1565 G; 2849 T; 0 other;

  Query Match             36.2%;  Score 630.8;  DB 22;  Length 8760;
  Best Local Similarity   99.7%;  Pred. No. 2.7e-173;
  Matches  632;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy   1110 AAATGCTTCGGACAAAGAAATCGTTTGGGTTATGCGAATCACAGTGTTTGTGTTTGGAGC 1169
           | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4220 ACAGGCTTCGGACAAAGAAATCGTTTGGGTTATGCGAATCACAGTGTTTGTGTTTGGAGC 4279

Qy   1170 ATCTGCAACAGCCATGGCCTTGCTGACGAAAACTGTGTATGGGCTCTGGTACCTCAGTTC 1229
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4280 ATCTGCAACAGCCATGGCCTTGCTGACGAAAACTGTGTATGGGCTCTGGTACCTCAGTTC 4339

Qy   1230 TGACCTTGTTTACATCGTTATCTTCCCCCAGCTGCTTTGTGTACTCTTTGTTAAGGGAAC 1289
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4340 TGACCTTGTTTACATCGTTATCTTCCCCCAGCTGCTTTGTGTACTCTTTGTTAAGGGAAC 4399

Qy   1290 CAACACCTATGGGGCCGTGGCAGGTTATGTTTCTGGCCTCTTCCTGAGAATAACTGGAGG 1349
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4400 CAACACCTATGGGGCCGTGGCAGGTTATGTTTCTGGCCTCTTCCTGAGAATAACTGGAGG 4459

Qy   1350 GGAGCCATATCTGTATCTTCAGCCCTTGATCTTCTACCCTGGCTATTACCCTGATGATAA 1409
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4460 GGAGCCATATCTGTATCTTCAGCCCTTGATCTTCTACCCTGGCTATTACCCTGATGATAA 4519

Qy   1410 TGGTATATATAATCAGAAATTTCCATTTAAAACACTTGCCATGGTTACATCATTCTTAAC 1469
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4520 TGGTATATATAATCAGAAATTTCCATTTAAAACACTTGCCATGGTTACATCATTCTTAAC 4579

Qy   1470 CAACATTTGCATCTCCTATCTAGCCAAGTATCTATTTGAAAGTGGAACCTTGCCACCTAA 1529
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4580 CAACATTTGCATCTCCTATCTAGCCAAGTATCTATTTGAAAGTGGAACCTTGCCACCTAA 4639

Qy   1530 ATTAGATGTATTTGATGCTGTTGTTGCAAGACACAGTGAAGAAAACATGGATAAGACAAT 1589
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4640 ATTAGATGTATTTGATGCTGTTGTTGCAAGACACAGTGAAGAAAACATGGATAAGACAAT 4699

Qy   1590 TCTTGTCAAAAATGAAAATATTAAATTAGATGAACTTGCACTTGTGAAGCCACGACAGAG 1649
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4700 TCTTGTCAAAAATGAAAATATTAAATTAGATGAACTTGCACTTGTGAAGCCACGACAGAG 4759

Qy   1650 CATGACCCTCAGCTCAACTTTCACCAATAAAGAGGCCTTCCTTGATGTTGATTCCAGTCC 1709
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4760 CATGACCCTCAGCTCAACTTTCACCAATAAAGAGGCCTTCCTTGATGTTGATTCCAGTCC 4819

Qy   1710 AGAAGGGTCTGGGACTGAAGATAATTTACAGTGA 1743
          |||||||||||||||||||||||||||||||||
Db   4820 AGAAGGGTCTGGGACTGAAGATAATTTACAGTGA 4853


RESULT 7
AAF81710
ID   AAF81710 standard; cDNA; 1731 BP.
XX
AC   AAF81710;
XX
DT   01-JUN-2001  (first entry)
XX
DE   C. elegans high affinity choline transporter protein encoding cDNA.
XX
KW   High affinity choline transporter; cho-1; Alzheimer's disease;
KW   diagnosis; ss.
XX
OS   Caenorhabditis elegans.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1731
FT                   /*tag= a
FT                   /product= "high affinity choline transporter"
XX
PN   WO200116315-A1.
XX
PD   08-MAR-2001.
XX
PF   18-AUG-2000; 2000WO-JP05545.
XX
PR   27-AUG-1999;   99JP-0240642.
PR   27-DEC-1999;   99JP-0368991.
XX
PA   (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI   Haga T,  Okuda T;
XX
DR   WPI; 2001-226688/23.
DR   P-PSDB; AAB74663.
XX
```

PT New rat and human spinal cord high affinity choline transporters,
PT useful in diagnosis of Alzheimer's disease and screening promoters as
PT drugs for treating Alzheimer's disease -
XX
PS Claim 3; Page 57-62; 90pp; Japanese.
XX
CC The present sequence encodes a Caenorhabditis elegans high affinity
CC choline transporter protein designated cho-1. The cho-1 protein has
CC nootropic and neuroprotective activities. The cho-1 polynucleotide and
CC protein can be used for the diagnosis of diseases related to the
CC expression of cho-1 by comparing the cho-1 polynucleotide sequence in a
CC sample to that of a control. Drug compositions containing the cho-1
CC protein or expression promoters or inhibitors of cho-1 are useful for
CC treating disorders characterised by abnormal levels of cho-1, such as
CC Alzheimer's disease.
XX
SQ Sequence 1731 BP; 428 A; 373 C; 427 G; 503 T; 0 other;

Query Match 20.9%; Score 363.8; DB 22; Length 1731;
Best Local Similarity 55.1%; Pred. No. 1.4e-95;
Matches 862; Conservative 0; Mismatches 637; Indels 66; Gaps 5;

```
Qy     19 GGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTGGTTGGAATATGGGCTGCC 78
Db     16 GGTATCGTGGCCATTGTGTTCTTCTACGTGCTCATTCTTGTCGTTGGAATATGGGCGGGT 75

Qy     79 TGGAGAACCAAAA-----------ACAGTGGCAGCGCAGAAGAGCGCAGCGAAGCCATC 126
Db     76 AGAAAATCGAAAAGTTCAAAAGAGCTTGAATCAGAAGCCGGCGCGGCGACGGAAGAGGTG 135

Qy    127 ATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCTACCTGG 186
Db    136 ATGTTAGCTGGGAGAAACATCGGAACTCTTGTCGGAATTTCACAATGACTGCCACGTGG 195

Qy    187 GTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTATGGCCTA 246
Db    196 GTTGGCGGTGCTTATATCAATGGAACCGCCGAGGCTCTGTATAATGGAGGT------CTC 249

Qy    247 GCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTGTTCTTT 306
Db    250 CTTGGATGTCAGGCTCCAGTTGGATATGCAATTTCCCTTGTTATGGGAGGACTACTTTTC 309

Qy    307 GCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAAATCTAT 366
Db    310 GCAAAGAAAATGCGAGAAGAAGGATATATTACAATGCTCGATCCTTTTCAGCACAAATAT 369

Qy    367 GGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTCTGGGCT 426
Db    370 GGCCAACGAATCGGTGGCTTGATGTATGTTCCAGCACTTCTTGGTGAAACATTCTGGACA 429

Qy    427 GCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGATATGCAC 486
Db    430 GCAGCCATTCTTTCGGCACTTGGTGCAACACTGTCGGTAATTCTTGGAATCGACATGAAT 489

Qy    487 ATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGGCTCTAT 546
Db    490 GCATCAGTGACCCTGTCGGCCTGTATTGCCGTATTCTACACATTCACCGGTGGATACTAT 549

Qy    547 TCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGGATCAGC 606
Db    550 GCAGTCGCGTACACTGACGTCGTTCAACTATTTTGCATTTTCGTCGGTTTGTGGGTTTGC 609

Qy    607 GTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTGCATGCC 666
Db    610 GTGCCGGCGGCTATGGTGCATGATGGTGCGAAGGATATTTCCAGGAATGCAG-------- 661

Qy    667 AAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGGCTTGAT 726
Db    662 -------GCGACTGGATTGGAGAGATTGGAGGATTCAAAGAAACATCTCTCTGGATTGAT 714

Qy    727 AGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGGGTTCTC 786
Db    715 TGCATGCTTCTCCTTGTCTTTGGAGGAATTCCATGGCAAGTGTACTTCCAAAGAGTTCTC 774

Qy    787 TCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGGTGCCTG 846
Db    775 TCCTCAAAAACTGCTCATGGAGCACAGACGTTGTCGTTTGTGGCGGGCGTCGGATGCATT 834

Qy    847 GTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGGAACCAG 906
Db    835 CTCATGGCGATTCCACCAGCGTTGATCGGTGCAATTGCCAGGAACACAGACTGGAGAATG 894

Qy    907 ACTGCATATGGGCTTCC----------------AGATCCCAAGACTACAGAAGAGGCA 948
Db    895 ACTGATTATTCCCCATGGAACAATGGAACTAAGGTCGAATCGATTCCACCGGATAAGAGA 954

Qy    949 GACATGATTTTACCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGT 1008
Db    955 AACATGGTGGTCCCGTTGGTATTCCAGTATCTTACGCCAAGATGGGTCGCCTTTATTGGA 1014

Qy   1009 CTTGGTGCAGTTTCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGT 1068
Db   1015 CTCGGCGCAGTGTCGGCTGCTGTAATGTCATCTGCAGATTCATCTGTACTATCAGCAGCA 1074

Qy   1069 TCCATGTTTGCACGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAA 1128
Db   1075 TCAATGTTTGCTCACAACATCTGGAAGCTCACAATTCGCCCTCACGCGTCTGAAAAAGAA 1134

Qy   1129 ATCGTTTGGGTTATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCC 1188
Db   1135 GTGATAATTGTGATGAGAATAGCCATCATCTGTGTTGGTATCATGGCAACCATCATGGCA 1194

Qy   1189 TTGCTGACGAAAACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTT 1248
Db   1195 CTTACCATTCAATCCATCTATGGGCTTTGGTATCTTTGTGCAGATTTGGTCTACGTCATA 1254

Qy   1249 ATCTTCCCCCAGCTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTG 1308
Db   1255 CTCTTCCCTCAACTATTATGTGTTGTATATATGCCACGTAGCAATACGTATGGCTCATTG 1314

Qy   1309 GCAGGTTATGTTTCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTT 1368
Db   1315 GCTGGCTATGCAGTCGGTCTTGTGCTCCGTTTGATTGGAGGCGAGCCACTTGTATCGCTG 1374

Qy   1369 CAGCCCTTGATCTTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAA 1428
Db   1375 CCAGCGTTCTTCCATTATCCAATGTATACGGATGGGG----------------TACAGTAT 1419

Qy   1429 TTTCCATTTAAAACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTAT 1488
Db   1420 TTCCCATTCAGGACAACTGCTATGTTATCTTCAATGGCTACTATCTACATTGTATCAATA 1479

Qy   1489 CTAGCCAAGTATCTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCT 1548
Db   1480 CAATCGGAGAAGCTGTTCAAATCGGGACGTTTGTCTCCGGAGTGGGACGTAATGGGTTGT 1539

Qy   1549 GTTGT 1553
Db   1540 GTAGT 1544
```

RESULT 8
AAD02461
ID AAD02461 standard; cDNA; 386 BP.
XX
AC AAD02461;
XX
DT 24-APR-2001 (first entry)
XX
DE Mouse P4P6B1 cDNA fragment.
XX
KW Mouse; OMA protein; obese mice adipocyte; P4P6B1;
KW fuel metabolism disorder; therapy; obesity; diabetes; gene therapy;
KW anorectic; antidiabetic; ss.
XX
OS Mus sp.

XX
PN WO200078950-A2.
XX
PD 28-DEC-2000.
XX
PF 13-JUN-2000; 2000WO-US16217.
XX
PR 22-JUN-1999; 99US-0141515.
XX
PA (AMYL-) AMYLIN PHARM INC.
XX
PI Sierzega M, Albrandt K;
XX
DR WPI; 2001-112322/12.
XX
PT Novel obese mice adipocyte polypeptides useful in diagnosis and
PT treatment of disorders of fuel metabolism such as obesity or diabetes
PT -
XX
PS Disclosure; Fig 2; 83pp; English.
XX
CC The present sequence is mouse P4P6B1 cDNA fragment. P4P6B1 cDNA encodes
CC OMA (obese mice adipocyte) protein. The P4P6B1 fragment was generated by
CC RNA fingerprinting using random primers P4 and P6.
CC OMA is used as a diagnostic reagent for diagnosing a disorder of fuel
CC metabolism in an underweight or an overweight individual, by detecting
CC the transcription level of a gene encoding OMA, which is induced or
CC repressed in an individual by a factor such as genetic obesity, fasting
CC and refeeding of a fasted individual. OMA is useful in the generation of
CC antibodies, for use in pharmaceutical compositions and for studying
CC DNA/protein interactions. Nucleic acids encoding OMA are involved in
CC gene therapy. An inhibitor of OMA or an antisense oligonucleotide that
CC inhibits expression of OMA are useful for treating disorders of fuel
CC metabolism such as obesity or diabetes.
XX
SQ Sequence 386 BP; 76 A; 109 C; 90 G; 111 T; 0 other;

Query Match 16.0%; Score 279.6; DB 22; Length 386;
Best Local Similarity 86.3%; Pred. No. 2.7e-71;
Matches 309; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

```
Qy    742 ATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGGGTTCTCTCTTCTTCCTCAGCC 801
          ||||||||||||||||||||||||||| ||||| |||||||| |||||||| |||||||||
Db     16 ATGCTGGGTGGAATCCCATGGCAAGCCTACTTCCAGAGGGTCCTCTCTTCATCCTCAGCC 75

Qy    802 ACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGGTGCCTGGTGATGGCCATCCCA 861
          |||||||||| || ||||||||||||||||||| |||||||||||||||||  | ||
Db     76 ACCTATGCTCAGGTACTGTCCTTCCTGGCAGCTTTTGGGTGCCTGGTGATGGCTCTACCC 135

Qy    862 GCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGGAACCAGACTGCATATGGGCTT 921
          ||||||  ||| || || |||||||| || ||||||||||||||||||| || |||  |
Db    136 GCCATATGCATAGGAGCTATTGGAGCTTCCACAGACTGGAACCAGACTGCCTACGGGTAT 195

Qy    922 CCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTACCAATTGTTCTGCAGTATCTC 981
          ||||||||||||||||  || || |||||||||||| | || || |||||||||||  |||
Db    196 CCAGATCCCAAGACTAAGGAGGAAGCAGACATGATTCTCCCGATCGTTCTGCAGTACCTC 255

Qy    982 TGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTTTCTGCTGCTGTTATGTCATCA 1041
          |||||||||| || || |||||||| |||||||| ||||| ||||||||| ||||| |||
Db    256 TGCCCTGTGTACATCTCCTTCTTTGGGCTTGGTGCTGTTTCAGCTGCTGTCATGTCCTCA 315

Qy   1042 GCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCACGGAACATCTACCAGCTTT 1099
          || || || |||||| |||| || ||||| |||||||| ||||| |||||||||| ||
Db    316 GCTGACTCGTCCATCCTGTCGGCGAGTTCTATGTTTGCTCGGAATATCTACCAGCATT 373
```

RESULT 9
ABL29569
ID ABL29569 standard; DNA; 1729 BP.
XX
AC ABL29569;
XX
DT 26-MAR-2002 (first entry)

XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 40180.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 40180; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1729 BP; 312 A; 498 C; 517 G; 402 T; 0 other;

Query Match 15.3%; Score 266.6; DB 23; Length 1729;
Best Local Similarity 52.9%; Pred. No. 3.9e-67;
Matches 819; Conservative 0; Mismatches 569; Indels 161; Gaps 5;

```
Qy      8 TCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTGGTTGGAA 67
          || || | |  ||  || |    ||| |  | |||||||| || ||  || | ||||| |
Db      5 TCAATATCGCTGGCGTGGTGAGCATCGTGCTCTTCTACCTCCTGATCCTGGTCGTTGGCA 64

Qy     68 TATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAAGCCATCA 127
          | ||||| |   | |  |   |  | |         | || |||          |   |||
Db     65 TTTGGGCCGGTCGCAAGAAGCAGTCCGGCAATGATTCGGAGGAG---------GAGGTCA 115

Qy    128 TAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCTACCTGGG 187
          |  | |  || ||    || ||  | || || ||    || |||||||| || |||||||
Db    116 TGCTGGCCGGACGCTCCATCGGCCTCTTCGTGGGCATCTTCACCATGACGGCCACCTGGG 175

Qy    188 TCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTATGGCCTAG 247
          | || || || || ||||| ||||| || || ||  | ||   | | |||        || |
Db    176 TGGGTGGCGGCTACATCAACGGCACGGCGGAGGCTATATACACATCGGGT------CTGG 229

Qy    248 CTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTGTTCTTTG 307
            |||   ||||| ||| |||||||  |||| ||  || | || ||||||| | |||||||
Db    230 TGTGGTGCCAGGCTCCATTTGGATACGCTCTAAGCTTGGTATTTGGTGGCATCTTCTTTG 289

Qy    308 CAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAAATCTATG 367
          | || || |||||      ||| ||  | |||||||| || ||||| ||| |   || ||
Db    290 CCAATCCCATGCGCAAGCAGGGTTACATCACCATGTTGGATCCGTTGCAGGATTCCTTTG 349

Qy    368 GAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTCTGGGCTG 427
          |  | || ||||| ||| | || ||  | || || ||    || ||  | || ||||| |
```

```
Db    350 GTGAGCGGATGGGAGGATTGCTCTTCCTGCCCGCTCTATGCGGTGAGGTCTTTTGGGCAG 409
Qy    428 CAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGATATGCACA 487
Db    410 CCGGCATCCTGGCTGCACTTGGCGCCACTCTATCGGTGATCATCGACATGGATCACCGCA 469
Qy    488 TTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGGCTCTATT 547
Db    470 CCTCGGTGATCCTGTCCTCCTGCATCGCCATCTTCTACACACTGTTCGGTGGACTGTACT 529
Qy    548 CTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGGATCAGCG 607
Db    530 CCGTGGCGTATACGGACGTGATCCAGTTGTTCTGCATCTTCATCGGTCTGTGGATGTGCA 589
Qy    608 TCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTGCATGCCA 667
Db    590 TTCCCTTCGCCTGGAGCAACGAGCACGTGGGCAGCCT---------------------- 626
Qy    668 AATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGGCTTGATA 727
Db    627 ------------------------------------------------------------ 626
Qy    728 GTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGGGTTCTCT 787
Db    627 ----------------------------------------------GAGTGACCT 635
Qy    788 CTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGGTGCCTGG 847
Db    636 GGAGCAAAACGGCAGGAAGGGCCCAGCTTCTGTCCTATGTTGCAGCCGCCGGATGCATTT 695
Qy    848 TGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGGAACCAGA 907
Db    696 TGATGGCCATTCCCCCGGTGCTCATCGGAGCGATTGCCAAGGCTACACCTTGGAACGAGA 755
Qy    908 CTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTACCAATTG 967
Db    756 CAGATTACAAGGGACCCTATCCCCTGACCGTGGACGAGACGAGCATGATTCTGCCCATGG 815
Qy    968 TTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTTTCTGCTG 1027
Db    816 TGCTGCAGTACCTCACGCCTGACTTCGTGTCCTTCTTTGGATTGGGCGCTGTTTCCGCCG 875
Qy   1028 CTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCACGGAACA 1087
Db    876 CCGTGATGTCCTCCGCCGACTCCTCGGTGCTCTCCGCCGCCTCCATGTTCGCTCGGAACG 935
Qy   1088 TCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTTATGCGAA 1147
Db    936 TGTACAAATTGATTTTCCGTCAGAAGGCGTCCGAGATGGAAATCATTTGGGTGATGCGAG 995
Qy   1148 TCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAAACTGTGT 1207
Db    996 TCGCCATCATTGTGGTGGGCATCCTGGCTACCATCATGGCCCTCACCATTCCCTCCATCT 1055
Qy   1208 ATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG---CTGC 1264
Db   1056 ACGGTTTGTGGTCCATGTGCTCGGATCTGGTCTACGTCATTCTGTTCCCGCAGCTACTGA 1115
Qy   1265 TTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTTTCTG 1324
Db   1116 TGGTGGTGCACTTCAAGAAGCACTGCAACACGTACGGCAGCCTGTCGGCATACATTGTGG 1175
Qy   1325 GCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATCTTCT 1384
Db   1176 CCCTGGCCATCCGACTGTCGGGCGGTGAGGCCATCTTGGGACTGGCTCCATTGATCAAGT 1235
Qy   1385 ACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAAACAC 1444
Db   1236 ATCCCGGCTACGACGAGGAGACCAAGG---------AGCAGATGTTCCCCTTCCGCACCA 1286
Qy   1445 TTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTATCTAT 1504
Db   1287 TGGCCATGCTGCTCAGCCTGGTCACGCTCATCTCGGTCTCCTGGTGGACTAAAATGATGT 1346

Qy   1505 TTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGT 1553
Db   1347 TTGAGTCCGGCAAGTTGCCGCCCAGCTACGACTACTTCCGCTGTGTGGT 1395

RESULT 10
ABK73210
ID   ABK73210 standard; DNA; 1461 BP.
XX
AC   ABK73210;
XX
DT   13-AUG-2002  (first entry)
XX
DE   Bacillus licheniformis genomic sequence tag (GST) #501.
XX
KW   Differential gene expression; genomic sequenced tag; GST;
KW   altered culture condition; environmental stress;
KW   physiological provocation; ds.
XX
OS   Bacillus licheniformis.
XX
PN   WO200229113-A2.
XX
PD   11-APR-2002.
XX
PF   05-OCT-2001; 2001WO-US31437.
XX
PR   06-OCT-2000; 2000US-0680598.
PR   27-MAR-2001; 2001US-279526P.
XX
PA   (NOVO ) NOVOZYMES BIOTECH INC.
PA   (NOVO ) NOVOZYMES AS.
XX
PI   Berka R,  Clausen IG;
XX
DR   WPI; 2002-416684/44.
XX
PT   Monitoring differential expression of several genes in first Bacillus
PT   cell relative to expression of same genes in one or more second
PT   Bacillus cells, by using substrate containing Bacillus genomic
PT   sequenced tag array -
XX
PS   Claim 4; SEQ ID NO 501; 200pp; English.
XX
CC   The invention describes a method of monitoring differential expression of
CC   genes in a first Bacillus cell relative to expression of the genes in
CC   other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC   isolated from Bacillus cells to a substrate containing array of Bacillus
CC   genomic sequenced tags (GST), examining the array, and determining
CC   relative gene expression by an observed hybridisation reporter signal of
CC   a spot in the array. The method is useful for measuring the expression of
CC   genes in a first Bacillus cell relative to expression of the same genes
CC   in one or more second Bacillus cells. The method is useful for monitoring
CC   global expression of several genes from a Bacillus cell, discovering new
CC   genes, identifying possible functions of unknown open reading frames and
CC   monitoring gene copy number variation and stability. Monitoring changes
CC   in expression of genes may be used to provide a representation of the way
CC   in which Bacillus cells adapt to changes in culture conditions,
CC   environmental stress or other physiological provocation. Extensive
CC   follow-up characterisation is unnecessary, when one spot on an array
CC   equals one gene or one open reading frame, since sequence information is
CC   available. This sequence represents a genomic sequence tag (GST) used in
CC   the method of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1461 BP; 315 A; 355 C; 382 G; 409 T; 0 other;

  Query Match           13.9%;  Score 242.6;  DB 24;  Length 1461;
  Best Local Similarity 52.4%;  Pred. No. 3.7e-60;
```

```
  Matches  663;  Conservative    0;  Mismatches  554;  Indels   48;  Gaps    4;

Qy     151 TTATTGGTTGGTGGATTTACCATGACAGCTACCTGGGTCGGAGGAGGGTATATCAATGGC 210
           || || || ||      | || ||| | || || |||||||| || || ||||| || ||
Db     125 TTTTTCGTCGGAATGGTGACGATGGCCGCAACATGGGTCGGCGGCGGATATATTAACGGA 184

Qy     211 ACAGCTGAAGCAGTTTATGTACCAGGTTATGGCCTAGCTTGGGCTCAGGCACCAATTGGA 270
           || || ||| |   |||          |  | |||||    ||||| || || ||    ||
Db     185 ACGGCCGAATCGACTTACA------GCGACGGCCTCATCTGGGCCCAAGCGCCTTGGGGC 238

Qy     271 TATTCTCTTAGTCTGATTTTAGGTGGCCTGTTCTTTGCAAAACCTATGCGTTCAAAGGGG 330
           ||  |  | || |||||| | || ||  | ||||| || | |   |||||      |
Db     239 TACGCATTGAGCCTGATTATCGGCGGTATTTTCTTCGCCAGAAAAATGCGCCGTCATCAA 298

Qy     331 TATGTGACCATGTTAGACCCGTTTCAGCAAATCTATGGAAAACGCATGGGCGGACTCCTG 390
           | | |||||||  | || ||  |  | ||   ||  || || || |||| |||  |  |
Db     299 TTTATGACCATTATCGATCCTCTCGAACAGCGCTTCGGTAAGCGGATGGCCGGCGTTTTA 358

Qy     391 TTTATTCCTGCACTGATGGGAGAAATGTTCTGGGCTGCAGCAATTTTCTCTGCTTTGGGA 450
           | ||| || || ||| | |||||| |||| |||   || || || ||  | || |||||
Db     359 TATATACCGGCGCTGTTAGGAGAATTGTTTTGGAGCGCCGCGATCTTAACGGCATTGGGC 418

Qy     451 GCCACCATCAGCGTGATCATCGATGTGGATATGCACATTTCTGTCATCATCTCTGCACTC 510
           | ||  || |  |||| |  || | ||| | || |  ||  | ||  | || ||  |
Db     419 ACGACTTTCGGAATGATTCTGAATATCGATTTCCAAACGTCGATTATTCTTTCGGCGATG 478

Qy     511 ATTGCCACTCTGTACACACTGGTGGGAGGGCTCTATTCTGTGGCCTACACTGATGTCGTT 570
           || ||||     ||  ||  |||  |||||  | |    | || || | ||| |||||| ||
Db     479 ATCGCCATCGCTTATACGGTGGCCGGAGGCATGTGGGCAGTTGCTTTCACAGATGTCTTT 538

Qy     571 CAGCTCTTTTGCATTTTTGTAGGGCTGTGGATCAGCGTCCCCTTTGCATTGTCACATCCT 630
           ||  |  ||  |||||  | |||||||      |   |||||| |||| |||||| ||
Db     539 CAAATGATTGTCATTTTGCTCGGGCTGTTTTTAGTCGTCCCATTTGTATTGTCGAATGTC 598

Qy     631 GCAG----------------------TCGCAGACATCGGGTTCACTGCTGTGCATGCCAA 668
           |  |                      |  ||| ||     |   | || | ||     |
Db     599 GGTGCTCTCGATAGCGTATGGGCAAATTACAGGCACGATTTCGGCAGCAGCGCCAATCTG 658

Qy     669 ATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCT-----TGGCTT 723
           | ||       |  |  ||||                ||| ||  | |     |||
Db     659 CTTCCGCCGTTGGACGGCTGGAAAAACCCGGATTGGGGAAACCTGTTTTGGAACTGGTGG 718

Qy     724 GATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGGGTT 783
           ||||  |   ||  | || ||  | || || ||| |||||||  |  |||||||| | |||
Db     719 GATAATGCGCTCCTCTTAATTTTCGGAGGTATCGCATGGCAGGTGTACTTTCAGCGCGTT 778

Qy     784 CTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGGTGC 843
           || ||  |    || |  | |   ||      || |||  |  | || |   |    |||
Db     779 CTTTCGGCAAAATCGGAAAGCGCCGCCATGTGGCAGTCGATAATTGCCGGAGTGATCTGC 838

Qy     844 CTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGGAAC 903
           |  |    ||||| ||   | || |||| || ||     |||      || || |||| |
Db     839 ATCATTGCCGCCATTCCGTGCGTAATCATCGGAGCTGCCGGAAACAGTACCGATTGGAGC 898

Qy     904 CAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTACCA 963
           | |   | |     |  ||| |||  |  | |                ||||||| ||
Db     899 CTGTTCGGAGCGAGCGCTCCGGATAACCCGGCG--------------ATGATTTTGCCG 943

Qy     964 ATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTTTCT 1023
           ||   |||| |    || |    ||  |   | | ||  | |||||| |  |
Db     944 CAAACGCTTGCGTATTTGACGCCAGGAATCATCGCAGGCCTCGGCTTGGGTGCAATCGCA 1003

Qy    1024 GCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCACGG 1083
           || || ||  ||||||          || || || ||  | || |||   || |||   ||| ||
Db    1004 GCAGCCGTCATGTCAAGCATGGACTCATCGATTCTATCGGCATCATCAATGGCCGCATGG 1063

Qy    1084 AACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTTATG 1143
           || || ||||  |     | |||  |  || ||  |  | |||  |  |          ||  |
Db    1064 AATATTTACCGTCCGCTCATCAAGCCGAAGGCCACCCAAAAACAGCTGCAAAAAGTCGTC 1123

Qy    1144 CGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAAACT 1203
           |  | || |  || |||| || ||    ||| | | ||| ||  ||          ||||
Db    1124 AAACGCTCAATCATTTTGTTCGGCGCGGGAGCAGCGGTCATCGCGCTGAATGTCAAAAGC 1183

Qy    1204 GTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAGCTG 1263
           || |||    | |||||  |   ||| ||  | |||||   | || | || ||||||  |
Db    1184 GTTTATACTTTATGGTATTTGGCTTCGGATTTAGTTTATTGCATTCTTTTTCCCCAGTTA 1243

Qy    1264 CTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTTTCT 1323
                 |  ||||||  |||  || | ||    || ||| |  | || || | ||    |
Db    1244 ACAATGGCCCTCTTTTATAAAAGAGCAAATCTTTACGGGTCGATTGCTGGATTTGCAGTT 1303

Qy    1324 GGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATCTTC 1383
           |   || | |||||  |    || || || ||     |     ||| |||  |  |
Db    1304 GCAGTCATTCTGAGGCTCGGCGGTGGTGAACCCGCATTCGGCATTCCGCCGCTTCTGCCG 1363

Qy    1384 TACCC 1388
           || ||
Db    1364 TATCC 1368


RESULT 11
AAH49201
ID   AAH49201 standard; DNA; 10140 BP.
XX
AC   AAH49201;
XX
DT   26-NOV-2001  (first entry)
XX
DE   Human CHOT promoter and exon 1 region DNA.
XX
KW   CHOT; human; choline transporter; chromosome 2q11-13; nootropic;
KW   neuroprotective; gene therapy; antisense therapy; degenerative disease;
KW   cognitive disorder; Alzheimer's disease; ds.
XX
OS   Homo sapiens.
XX
PN   DE10009055-A1.
XX
PD   30-AUG-2001.
XX
PF   28-FEB-2000; 2000DE-1009055.
XX
PR   28-FEB-2000; 2000DE-1009055.
XX
PA   (BRUE/) BRUESS M.
PA   (BOEN/) BOENISCH H.
XX
PI   Bruess M,  Boenisch H;
XX
DR   WPI; 2001-590709/67.
XX
PT   A new gene encoding human choline transporter, designated hCHOT is
PT   located on chromosome 2q11-13 and is useful to treat degenerative
PT   disorders such as Alzheimer's disease -
XX
PS   Disclosure; Page 4-7; 12pp; German.
XX
CC   This invention describes a novel gene encoding human choline transporter,
CC   designated hCHOT which is located on chromosome 2q11-13. The products of
CC   the invention have nootropic and neuroprotective activity and can be used
CC   for gene or antisense therapy. (I) is used to treat degenerative disease,
CC   particularly cognitive disorders such as Alzheimer's disease. Sense and
CC   antisense oligonucleotides derived from the gene may be used in
CC   diagnostics and other techniques. This sequence represents the promoter
CC   region and the exon 1 fragment encoding the human CHOT protein described
CC   in the invention.
XX
SQ   Sequence 10140 BP; 3296 A; 2001 C; 2030 G; 2813 T; 0 other;

  Query Match             10.4%;  Score 180.8;  DB 22;  Length 10140;
  Best Local Similarity   98.9%;  Pred. No. 1.3e-41;
  Matches  182;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ATGGCTTTCCATGTGGAAGCACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    9892 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 9951

Qy      61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9952 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 10011

Qy     121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   10012 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGGT 10071

Qy     181 ACCT 184
           || |
Db   10072 ACGT 10075
```

RESULT 12
ABL29568/c
ID ABL29568 standard; DNA; 4223 BP.
XX
AC ABL29568;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 40177.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 40177; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4223 BP; 1128 A; 1024 C; 964 G; 1107 T; 0 other;

Query Match 10.3%; Score 179.6; DB 23; Length 4223;
Best Local Similarity 57.2%; Pred. No. 1.8e-41;
Matches 357; Conservative 0; Mismatches 249; Indels 18; Gaps 1;

```
Qy     260 CACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTGTTCTTTGCAAAACCTATGC 319
           ||  || ||  | |   ||   ||  |  ||||||||| | ||||||||| || || ||||
Db    2907 CAAAAAGTGATTCTATCCTCCTTCCAAACTTAGGTGGCATCTTCTTTGCCAATCCCATGC 2848

Qy     320 GTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAAATCTATGGAAAACGCATGG 379
           |    |||| ||  | ||||||||| || ||||| ||| |   || |||  | || ||||
Db    2847 GCAAGCAGGGTTACATCACCATGTTGGATCCGTTGCAGGATTCCTTTGGTGAGCGGATGG 2788

Qy     380 GCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTCTGGGCTGCAGCAATTTTCT 439
           | ||| | || ||  | || || ||    || ||  | || ||||| || |  ||  |
Db    2787 GAGGATTGCTCTTCCTGCCCGCTCTATGCGGTGAGGTCTTTTGGGCAGCCGGCATCCTGG 2728

Qy     440 CTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGATATGCACATTTCTGTCATCA 499
           ||||  | ||  |||||  |    ||||||||||  |||||   | ||  || || |||
Db    2727 CTGCACTTGGCGCCACTCTATCGGTGATCATCGACATGGATCACCGCACCTCGGTGATCC 2668

Qy     500 TCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGGCTCTATTCTGTGGCCTACA 559
           | ||  |   ||| ||||   | |||||||| | || || || || || ||||| || |
Db    2667 TGTCCTCCTGCATCGCCATCTTCTACACACTGTTCGGTGGACTGTACTCCGTGGCGTATA 2608

Qy     560 CTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGGATCAGCGTCCCCTTTGCAT 619
           | || ||  | ||| | || ||||| ||  | || ||||||||  || | ||||| || |
Db    2607 CGGACGTGATCCAGTTGTTCTGCATCTTCATCGGTCTGTGGATGTGCATTCCCTTCGCCT 2548

Qy     620 TGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTGCATGCCAAATACCAAAAGC 679
           |   |       || |   | |   |   |  ||| ||                      
Db    2547 GGAGCAACGAGCACGTGGGCAGCCTGAGTGACCTGGAGGTGGAT---------------- 2504

Qy     680 CGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGGCTTGATAGTTTTCTGTTGT 739
             ||| | ||    || ||  |   |     |   | ||         || ||
Db    2503 --TGGATTGGGCACGTGGAGCCTAAAAAGCATTGGCTGTACATAGACTACGGCTTGCTGC 2446

Qy     740 TGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGGGTTCTCTCTTCTTCCTCAG 799
           |  |  | ||||| || || ||||| |  ||||| ||| |||| || ||          | |
Db    2445 TCGTCTTTGGTGGCATTCCCTGGCAGGTCTACTTCCAGCGGGTGCTGTCCAGCAAAACGG 2386

Qy     800 CCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGGTGCCTGGTGATGGCCATCC 859
           |       || ||  | |||||||   | |||||   ||| ||| |  ||||||||| |
Db    2385 CAGGAAGGGCCCAGCTTCTGTCCTATGTTGCAGCCGCCGGATGCATTTTGATGGCCATTC 2326

Qy     860 CAGCCATACTCATTGGGGCCATTG 883
           |  |  | ||||| || || ||||
Db    2325 CCCCGGTGCTCATCGGAGCGATTG 2302

RESULT 13
AAS82193
ID  AAS82193 standard; cDNA; 1094 BP.
XX
AC  AAS82193;
XX
DT  13-FEB-2002 (first entry)
XX
DE  DNA encoding novel human diagnostic protein #17997.
XX
KW  Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW  food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS  Homo sapiens.
XX
PN  WO200175067-A2.
XX
PD  11-OCT-2001.
XX
PF  30-MAR-2001; 2001WO-US08631.
XX
PR  31-MAR-2000; 2000US-0540217.
PR  23-AUG-2000; 2000US-0649167.
XX
PA  (HYSE-) HYSEQ INC.
XX
PI  Drmanac RT, Liu C, Tang YT;
XX
DR  WPI; 2001-639362/73.
DR  P-PSDB; ABG18006.
XX
PT  New isolated polynucleotide and encoded polypeptides, useful in
```

```
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 17997; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1094 BP; 325 A; 210 C; 267 G; 292 T; 0 other;

  Query Match             9.6%;  Score 167.6;  DB 23;  Length 1094;
  Best Local Similarity   96.8%;  Pred. No. 2.8e-38;
  Matches  182;  Conservative    0;  Mismatches    4;  Indels    2;  Gaps    1;

Qy       1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60

Qy      61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||  |||||
Db      61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCG--GCGAA 118

Qy     121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     119 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGAT 178

Qy     181 ACCTGGGT 188
           || |  ||
Db     179 ACATATGT 186


RESULT 14
AAH49203
ID   AAH49203 standard; DNA; 240 BP.
XX
AC   AAH49203;
XX
DT   26-NOV-2001  (first entry)
XX
DE   Human CHOT exon 3 region DNA.
XX
KW   CHOT; human; choline transporter; chromosome 2q11-13; nootropic;
KW   neuroprotective; gene therapy; antisense therapy; degenerative disease;
KW   cognitive disorder; Alzheimer's disease; ds.
XX
OS   Homo sapiens.
XX
PN   DE10009055-A1.
XX
PD   30-AUG-2001.
XX
PF   28-FEB-2000; 2000DE-1009055.
XX
PR   28-FEB-2000; 2000DE-1009055.
XX
PA   (BRUE/) BRUESS M.
PA   (BOEN/) BOENISCH H.
XX
PI   Bruess M,  Boenisch H;
XX
DR   WPI; 2001-590709/67.
XX
PT   A new gene encoding human choline transporter, designated hCHOT is
PT   located on chromosome 2q11-13 and is useful to treat degenerative
PT   disorders such as Alzheimer's disease -
XX
PS   Disclosure; Page 7; 12pp; German.
XX
CC   This invention describes a novel gene encoding human choline transporter,
CC   designated hCHOT which is located on chromosome 2q11-13. The products of
CC   the invention have nootropic and neuroprotective activity and can be used
CC   for gene or antisense therapy. (I) is used to treat degenerative disease,
CC   particularly cognitive disorders such as Alzheimer's disease. Sense and
CC   antisense oligonucleotides derived from the gene may be used in
CC   diagnostics and other techniques. This sequence represents the exon 3
CC   fragment encoding the human CHOT protein described in the invention.
XX
SQ   Sequence 240 BP; 60 A; 45 C; 53 G; 82 T; 0 other;

  Query Match             9.4%;  Score 163;  DB 22;  Length 240;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-37;
  Matches  163;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     287 TTTTAGGTGGCCTGTTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAG 346
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60 TTTTAGGTGGCCTGTTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAG 119

Qy     347 ACCCGTTTCAGCAAATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGA 406
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 ACCCGTTTCAGCAAATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGA 179

Qy     407 TGGGAGAAATGTTCTGGGCTGCAGCAATTTTCTCTGCTTTGGG 449
          |||||||||||||||||||||||||||||||||||||||||||
Db     180 TGGGAGAAATGTTCTGGGCTGCAGCAATTTTCTCTGCTTTGGG 222


RESULT 15
ABA53620
ID   ABA53620 standard; DNA; 455 BP.
XX
AC   ABA53620;
XX
DT   01-FEB-2002  (first entry)
XX
DE   Human foetal liver single exon nucleic acid probe #1925.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00669.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
```

PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 1; SEQ ID NO 1925; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 455 BP; 121 A; 108 C; 85 G; 141 T; 0 other;

Query Match 8.9%; Score 155; DB 22; Length 455;
Best Local Similarity 100.0%; Pred. No. 8.3e-35;
Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    741 GATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGGGTTCTCTCTTCTTCCTCAGC 800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    266 GATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGGGTTCTCTCTTCTTCCTCAGC 325

Qy    801 CACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGGTGCCTGGTGATGGCCATCCC 860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    326 CACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGGTGCCTGGTGATGGCCATCCC 385

Qy    861 AGCCATACTCATTGGGGCCATTGGAGCATCAACAG 895
          |||||||||||||||||||||||||||||||||||
Db    386 AGCCATACTCATTGGGGCCATTGGAGCATCAACAG 420
```

Search completed: March 31, 2003, 16:33:42
Job time : 436 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 31, 2003, 15:57:11 ; Search time 4568 Seconds
(without alignments)
11104.683 Million cell updates/sec

Title: US-10-069-541-5
Perfect score: 1743
Sequence: 1 atggctttccatgtggaagg.....ctgaagataatttacagtga 1743

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1743	100.0	1743	6	BD012719	BD012719 High-affi
	2	1743	100.0	1743	9	AF276871	AF276871 Homo sapi
	3	1743	100.0	1743	23	BD005267	Bd005267 High-affi
	4	1743	100.0	1813	9	HSA401466	AJ401466 Homo sapi
	5	1743	100.0	5158	9	AB043997	AB043997 Homo sapi
	6	1394.2	80.0	1743	6	BD012718	BD012718 High-affi
	7	1394.2	80.0	1743	23	BD005266	Bd005266 High-affi
	8	1394.2	80.0	4904	10	AB030947	AB030947 Rattus no
	9	1375	78.9	1743	10	AF276872	AF276872 Mus muscu
	10	1373.4	78.8	1743	6	BD012720	BD012720 High-affi
	11	1373.4	78.8	1743	23	BD005268	Bd005268 High-affi
	12	1373.4	78.8	4938	6	AX080443	AX080443 Sequence
	13	1367	78.4	1743	10	MMU401467	AJ401467 Mus muscu
	14	867	49.7	2528	5	TMA420808	AJ420808 Torpedo m
	15	630.8	36.2	2239	9	HSA308384	AJ308384 Homo sapi
	16	630.8	36.2	190043	9	AC009963	AC009963 Homo sapi
c	17	502.8	28.8	174779	2	AC106657	AC106657 Rattus no
c	18	501.2	28.8	155131	2	AC102873	AC102873 Mus muscu
	19	431.4	24.8	3326	3	AY011119	AY011119 Limulus p
	20	405.8	23.3	3255	3	AY047521	AY047521 Drosophil
	21	363.8	20.9	1731	6	BD012717	BD012717 High-affi
	22	363.8	20.9	1731	23	BD005265	Bd005265 High-affi
	23	363.8	20.9	1985	3	AB030946	AB030946 Caenorhab
	24	279.6	16.0	386	6	AX080449	AX080449 Sequence
	25	242.6	13.9	1461	6	AX432086	AX432086 Sequence
	26	226	13.0	1657	9	HSA308383	AJ308383 Homo sapi
	27	179.6	10.3	1178	9	HSA308378	AJ308378 Homo sapi
	28	179.6	10.3	186989	3	AC007812	AC007812 Drosophil
	29	179.6	10.3	189117	3	AC009395	AC009395 Drosophil
	30	179.6	10.3	255624	3	AE003723	AE003723 Drosophil
c	31	167.6	9.6	140156	2	AC017381	AC017381 Drosophil
	32	163	9.4	2326	9	HSA308379	AJ308379 Homo sapi
	33	155	8.9	1467	9	HSA308382	AJ308382 Homo sapi
c	34	151.8	8.7	40893	3	CBRG45E19	AC084631 Caenorhab
	35	150.2	8.6	736	9	HSA308381	AJ308381 Homo sapi
	36	150	8.6	1308	9	HSA308380	AJ308380 Homo sapi
c	37	141.6	8.1	39908	3	CEC48D1	Z81049 Caenorhabdi
c	38	141.6	8.1	330724	2	CEY67H2	AL022475 Caenorhab
c	39	132	7.6	152021	2	AC010923	AC010923 Drosophil
	40	109	6.3	174779	2	AC106657	AC106657 Rattus no
c	41	51.8	3.0	10732	6	E32986	E32986 Gene encodi
	42	48.2	2.8	2781	1	BSU92466	U92466 Bacillus su
c	43	48.2	2.8	3933	1	AF011545	AF011545 Bacillus
c	44	48.2	2.8	213190	1	BSUB0004	Z99107 Bacillus su
	45	48	2.8	2426	4	SSC487855	AJ487855 Sus scrof

ALIGNMENTS

RESULT 1
BD012719
LOCUS BD012719 1743 bp DNA linear PAT 02-AUG-2002
DEFINITION High-affinity choline transporter.
ACCESSION BD012719
VERSION BD012719.1 GI:22092908
KEYWORDS WO 0116315-A/3.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1743)
AUTHORS Haga,T. and Okuda,T.
TITLE High-affinity choline transporter
JOURNAL Patent: WO 0116315-A 3 08-MAR-2001;

```
            JAPAN SCIENCE AND TECHNOLOGY CORP,TATSUYA HAGA,TAKASHI OKUDA
COMMENT     OS   Homo sapiens (human)
            PN   WO 0116315-A/3
            PD   08-MAR-2001
            PF   18-AUG-2000 WO 2000JP005545
            PR   27-AUG-1999 JP 99P   240642,27-DEC-1999 JP 99P   368991 PI
            TATSUYA HAGA,TAKASHI OKUDA
            PC   C12N15/12,C07K14/47,C12Q1/68,C07K19/00,C07K16/18,C12N5/10, PC
             A61K38/17,
            PC   A61K45/00,A61P25/28,G01N33/53,A01K67/027
            CC
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(1743).
FEATURES              Location/Qualifiers
     source           1. .1743
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
BASE COUNT      412 a    393 c    406 g    532 t
ORIGIN

Query Match           100.0%;  Score 1743;  DB 6;  Length 1743;
Best Local Similarity  100.0%;  Pred. No. 0;
Matches 1743;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60

Qy      61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120

Qy     121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180

Qy     181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240

Qy     241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300

Qy     301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360

Qy     361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420

Qy     421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 480

Qy     481 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 540

Qy     541 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 600

Qy     601 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 660

Qy     661 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 720

Qy     721 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 780

Qy     781 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 840

Qy     841 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 900

Qy     901 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 960

Qy     961 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1020

Qy    1021 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1080

Qy    1081 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140

Qy    1141 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1200

Qy    1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1260

Qy    1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1320

Qy    1321 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1380

Qy    1381 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1440

Qy    1441 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1500

Qy    1501 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1560

Qy    1561 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1620

Qy    1621 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1680

Qy    1681 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1740

Qy    1741 TGA 1743
             |||
Db    1741 TGA 1743


RESULT 2
```

```
AF276871
LOCUS       AF276871                1743 bp    mRNA    linear   PRI 27-NOV-2000
DEFINITION  Homo sapiens high affinity choline transporter (SLC5A7) mRNA,
            complete cds.
ACCESSION   AF276871
VERSION     AF276871.1  GI:10998441
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1743)
  AUTHORS   Apparsundaram,S., Ferguson,S.M., George,A.L. Jr. and Blakely,R.D.
  TITLE     Molecular cloning of a human, hemicholinium-3-sensitive choline
            transporter
  JOURNAL   Biochem. Biophys. Res. Commun. 276 (3), 862-867 (2000)
  MEDLINE   20483599
   PUBMED   11027560
REFERENCE   2  (bases 1 to 1743)
  AUTHORS   Apparsundaram,S., Ferguson,S.M. and Blakely,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-2000) Department of Pharmacology and Center for
            Molecular Neuroscience, Vanderbilt University, 23rd Avenue South at
            Pierce, Nashville, TN 37232-6420, USA
FEATURES             Location/Qualifiers
     source          1. .1743
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q12"
     gene            1. .1743
                     /gene="SLC5A7"
     CDS             1. .1743
                     /gene="SLC5A7"
                     /note="hCHT; solute carrier family 5 member 7"
                     /codon_start=1
                     /product="high affinity choline transporter"
                     /protein_id="AAG25940.1"
                     /db_xref="GI:10998442"
                     /translation="MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIV
                     GGRDIGLLVGGFTMTATWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFF
                     AKPMRSKGYVTMLDPFQQIYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVD
                     MHISVIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFT
                     AVHAKYQKPWLGTVDSSEVYSWLDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFL
                     AAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKTTEEADMILPIVLQYLCPVYISF
                     FGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITVFVFGASA
                     TAMALLTKTVYGLWYLSSDLVYIVIFPQLLCVLFVKGTNTYGAVAGYVSGLFLRITGG
                     EPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLP
                     PKLDVFDAVVARHSEENMDKTILVKNENIKLDELALVKPRQSMTLSSTFTNKEAFLDV
                     DSSPEGSGTEDNLQ"
BASE COUNT      412 a    393 c    406 g    532 t
ORIGIN
```

Query Match 100.0%; Score 1743; DB 9; Length 1743;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1743; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60
Db       1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60

Qy      61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120
Db      61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120

Qy     121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180
Db     121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180

Qy     181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240
Db     181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240

Qy     241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300
Db     241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300

Qy     301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360
Db     301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360

Qy     361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420
Db     361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420

Qy     421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 480
Db     421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 480

Qy     481 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 540
Db     481 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 540

Qy     541 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 600
Db     541 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 600

Qy     601 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 660
Db     601 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 660

Qy     661 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 720
Db     661 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 720

Qy     721 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 780
Db     721 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 780

Qy     781 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 840
Db     781 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 840

Qy     841 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 900
Db     841 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 900

Qy     901 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 960
Db     901 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 960

Qy     961 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1020
Db     961 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1020

Qy    1021 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1080
Db    1021 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1080

Qy    1081 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140
Db    1081 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140

Qy    1141 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1200
Db    1141 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1200

Qy    1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1260
Db    1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1260

Qy    1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1320
Db    1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1320

Qy    1321 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1380
```

```
Db    1321 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1380

Qy    1381 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1440

Qy    1441 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1500

Qy    1501 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1560

Qy    1561 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1620

Qy    1621 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1680

Qy    1681 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1740

Qy    1741 TGA 1743
           |||
Db    1741 TGA 1743
```

```
RESULT 3
BD005267
ID   BD005267   standard; DNA; HUM; 1743 BP.
XX
AC   BD005267;
XX
SV   BD005267.1
XX
DT   08-FEB-2002 (Rel. 70, Created)
DT   08-FEB-2002 (Rel. 70, Last updated, Version 1)
XX
DE   High-affinity choline transporter.
XX
KW   JP 03074455-T/3.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-1743
RA   Haga T., Okuda T.;
RT   "High-affinity choline transporter";
RL   Patent number JP03074455-T/3, 19-JAN-2001.
RL   JAPAN SCIENCE AND TECHNOLOGY CORP,TATSUYA HAGA,TAKASHI OKUDA.
XX
CC   OS   Homo sapiens (human)
CC   PN   JP 03074455-T/3
CC   PD   19-JAN-2001
CC   PF   18-AUG-2000 JP 2000005545
CC   PR   27-AUG-1999 JP 99P   240642,27-DEC-1999 JP 99P   368991
CC   PI   TATSUYA HAGA,TAKASHI OKUDA
CC   PC   C12N15/12,C07K14/47,C12Q1/68,C07K19/00,C07K16/18,C12N5/10,
CC   PC   A61K38/17,
CC   PC   A61K45/00,A61P25/28,G01N33/53,A01K67/027
CC   CC
CC   FH   Key             Location/Qualifiers
CC   FT   CDS             (1). .(1743)
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .1743
FT                   /db_xref="taxon:9606"
FT                   /organism="Homo sapiens"
XX
SQ   Sequence 1743 BP; 412 A; 393 C; 406 G; 532 T; 0 other;

  Query Match             100.0%;  Score 1743;  DB 23;  Length 1743;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1743;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60

Qy      61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120

Qy     121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180

Qy     181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240

Qy     241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300

Qy     301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360

Qy     361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420

Qy     421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 480

Qy     481 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 540

Qy     541 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 600

Qy     601 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 660

Qy     661 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 720

Qy     721 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 780

Qy     781 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 840

Qy     841 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 900

Qy     901 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 960
```

```
Qy     961 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1020

Qy    1021 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1080

Qy    1081 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140

Qy    1141 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1200

Qy    1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1260

Qy    1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1320

Qy    1321 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1380

Qy    1381 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1440

Qy    1441 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1500

Qy    1501 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1560

Qy    1561 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1620

Qy    1621 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1680

Qy    1681 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1740

Qy    1741 TGA 1743
           |||
Db    1741 TGA 1743
```

RESULT 4
HSA401466

```
LOCUS       HSA401466               1813 bp    mRNA    linear   PRI 16-AUG-2000
DEFINITION  Homo sapiens mRNA for high affinity choline transporter (CHT1
            gene).
ACCESSION   AJ401466
VERSION     AJ401466.1  GI:9843753
KEYWORDS    ChT1 gene; high affinity choline transporter.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1813)
  AUTHORS   Wieland,A., Bonisch,H. and Bruss,M.
  TITLE     Molecular cloning of the human and murine high affinity choline
            transporters and characterization ofthe human gene-structure
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1813)
  AUTHORS   Bruess,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-2000) Bruess M., University of Bonn, Pharmacology
            and Toxicology, Reuter str. 2b, D-53113 Bonn, GERMANY
FEATURES             Location/Qualifiers
     source          1. .1813
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q11-13"
                     /tissue_type="hypothalamus"
     gene            19. .1761
                     /gene="CHT1"
     CDS             19. .1761
                     /gene="CHT1"
                     /function="sodium- and chloride-dependent reuptake of
                     choline"
                     /codon_start=1
                     /evidence=experimental
                     /product="high affinity choline transporter"
                     /protein_id="CAC03717.1"
                     /db_xref="GI:9843754"
                     /translation="MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIV
                     GGRDIGLLVGGFTMTATWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFF
                     AKPMRSKGYVTMLDPFQQIYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVD
                     MHISVIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFT
                     AVHAKYQKPWLGTVDSSEVYSWLDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFL
                     AAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKTTEEADMILPIVLQYLCPVYISF
                     FGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITVFVFGASA
                     TAMALLTKTVYGLWYLSSDLVYIVIFPQLLCVLFVKGTNTYGAVAGYVSGLFLRITGG
                     EPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLP
                     PKLDVFDAVVARHSEENMDKTILVKNENIKLDELALVKPRQSMTLSSTFTNKEAFLDV
                     DSSPEGSGTEDNLQ"
BASE COUNT      440 a    406 c    417 g    550 t
ORIGIN
```

```
Query Match           100.0%;  Score 1743;  DB 9;  Length 1813;
Best Local Similarity  100.0%;  Pred. No. 0;
Matches 1743;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 78

Qy      61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 138

Qy     121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 198

Qy     181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 258

Qy     241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 318

Qy     301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 378

Qy     361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     379 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 438

Qy     421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    439 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 498
Qy    481 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 540
Db    499 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 558
Qy    541 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 600
Db    559 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 618
Qy    601 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 660
Db    619 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 678
Qy    661 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 720
Db    679 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 738
Qy    721 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 780
Db    739 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 798
Qy    781 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 840
Db    799 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 858
Qy    841 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 900
Db    859 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 918
Qy    901 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 960
Db    919 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 978
Qy    961 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1020
Db    979 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1038
Qy   1021 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1080
Db   1039 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1098
Qy   1081 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140
Db   1099 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1158
Qy   1141 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1200
Db   1159 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1218
Qy   1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1260
Db   1219 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1278
Qy   1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1320
Db   1279 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1338
Qy   1321 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1380
Db   1339 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1398
Qy   1381 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1440
Db   1399 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1458
Qy   1441 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1500
Db   1459 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1518
Qy   1501 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1560
Db   1519 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1578
```

```
Qy   1561 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1620
Db   1579 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1638
Qy   1621 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1680
Db   1639 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1698
Qy   1681 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1740
Db   1699 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1758
Qy   1741 TGA 1743
Db   1759 TGA 1761
```

RESULT 5
AB043997

```
LOCUS       AB043997                5158 bp    mRNA    linear   PRI 19-NOV-2000
DEFINITION  Homo sapiens mRNA for high-affinity choline transporter CHT1,
            complete cds.
ACCESSION   AB043997
VERSION     AB043997.1  GI:11231080
KEYWORDS    .
SOURCE      Homo sapiens spinal cord cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Okuda,T. and Haga,T.
  TITLE     Functional characterization of the human high-affinity choline
            transporter
  JOURNAL   FEBS Lett. 484 (2), 92-97 (2000)
  MEDLINE   20521663
REFERENCE   2  (bases 1 to 5158)
  AUTHORS   Okuda,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAY-2000) Takashi Okuda, University of Tokyo, Faculty
            of Medicine, Department of Neurochemistry; 7-3-1 Hongo, Bunkyo-ku,
            Tokyo 1130033, Japan (E-mail:okuda@m.u-tokyo.ac.jp,
            URL:http://park.ecc.u-tokyo.ac.jp/neurochemistry,
            Tel:81-3-5841-3560, Fax:81-3-6814-8154)
FEATURES             Location/Qualifiers
     source          1. .5158
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="spinal cord"
     CDS             277. .2019
                     /codon_start=1
                     /product="high-affinity choline transporter CHT1"
                     /protein_id="BAB18161.1"
                     /db_xref="GI:11231081"
                     /translation="MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIV
                     GGRDIGLLVGGFTMTATWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFF
                     AKPMRSKGYVTMLDPFQQIYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVD
                     MHISVIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFT
                     AVHAKYQKPWLGTVDSSEVYSWLDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFL
                     AAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKTTEEADMILPIVLQYLCPVYISF
                     FGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITVFVFGASA
                     TAMALLTKTVYGLWYLSSDLVYIVIFPQLLCVLFVKGTNTYGAVAGYVSGLFLRITGG
                     EPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLP
                     PKLDVFDAVVARHSEENMDKTILVKNENIKLDELALVKPRQSMTLSSTFTNKEAFLDV
                     DSSPEGSGTEDNLQ"
BASE COUNT     1613 a    1019 c     981 g    1545 t
ORIGIN
```

```
Query Match             100.0%;  Score 1743;  DB 9;  Length 5158;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 1743;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60
```

```
Db    277 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 336
Qy     61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120
Db    337 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 396
Qy    121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180
Db    397 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 456
Qy    181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240
Db    457 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 516
Qy    241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300
Db    517 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 576
Qy    301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360
Db    577 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 636
Qy    361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420
Db    637 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 696
Qy    421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 480
Db    697 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 756
Qy    481 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 540
Db    757 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 816
Qy    541 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 600
Db    817 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 876
Qy    601 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 660
Db    877 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 936
Qy    661 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 720
Db    937 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 996
Qy    721 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 780
Db    997 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 1056
Qy    781 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 840
Db   1057 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 1116
Qy    841 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 900
Db   1117 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 1176
Qy    901 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 960
Db   1177 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 1236
Qy    961 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1020
Db   1237 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1296
Qy   1021 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1080
Db   1297 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1356
Qy   1081 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140
Db   1357 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1416
Qy   1141 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1200
Db   1417 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1476
Qy   1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1260
Db   1477 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1536
Qy   1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1320
Db   1537 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1596
Qy   1321 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1380
Db   1597 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1656
Qy   1381 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1440
Db   1657 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1716
Qy   1441 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1500
Db   1717 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1776
Qy   1501 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1560
Db   1777 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1836
Qy   1561 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1620
Db   1837 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1896
Qy   1621 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1680
Db   1897 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1956
Qy   1681 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1740
Db   1957 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 2016
Qy   1741 TGA 1743
Db   2017 TGA 2019
```

RESULT 6
BD012718

LOCUS	BD012718 1743 bp DNA linear PAT 02-AUG-2002
DEFINITION	High-affinity choline transporter.
ACCESSION	BD012718
VERSION	BD012718.1 GI:22092907
KEYWORDS	WO 0116315-A/2.
SOURCE	Rattus norvegicus.
ORGANISM	Rattus norvegicus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE	1 (bases 1 to 1743)
AUTHORS	Haga,T. and Okuda,T.
TITLE	High-affinity choline transporter
JOURNAL	Patent: WO 0116315-A 2 08-MAR-2001; JAPAN SCIENCE AND TECHNOLOGY CORP,TATSUYA HAGA,TAKASHI OKUDA
COMMENT	OS Rattus norvegicus (rat) PN WO 0116315-A/2 PD 08-MAR-2001 PF 18-AUG-2000 WO 2000JP005545 PR 27-AUG-1999 JP 99P 240642,27-DEC-1999 JP 99P 368991 PI TATSUYA HAGA,TAKASHI OKUDA PC C12N15/12,C07K14/47,C12Q1/68,C07K19/00,C07K16/18,C12N5/10, PC A61K38/17, PC A61K45/00,A61P25/28,G01N33/53,A01K67/027

```
            CC
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(1743).
FEATURES             Location/Qualifiers
     source          1. .1743
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
BASE COUNT       414 a    402 c    404 g    523 t
ORIGIN
```

Query Match 80.0%; Score 1394.2; DB 6; Length 1743;
Best Local Similarity 87.5%; Pred. No. 0;
Matches 1525; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

```
Qy       1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60
Db       1 ATGCCTTTCCATGTAGAAGGACTAGTAGCGATTATCCTGTTCTACCTTCTTATATTTCTG 60

Qy      61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120
Db      61 GTTGGAATATGGGCTGCATGGAAAACCAAAAACAGCGGTAATGCAGAAGAACGCAGCGAA 120

Qy     121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180
Db     121 GCCATCATAGTTGGGGGCCGAGACATTGGTTTGTTGGTTGGTGGTTTTACCATGACAGCC 180

Qy     181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240
Db     181 ACCTGGGTTGGAGGAGGTTACATCAACGGGACAGCTGAAGCAGTTTATGGGCCAGGTTGT 240

Qy     241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300
Db     241 GGTCTAGCTTGGGCTCAGGCACCCATTGGATATTCTCTGAGTCTGATTTTAGGTGGCCTG 300

Qy     301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360
Db     301 TTTTTTGCAAAACCTATGCGTTCCAAGGGATATGTGACTATGTTAGACCCGTTTCAACAG 360

Qy     361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420
Db     361 ATCTATGGAAAGCGCATGGGTGGGCTGCTGTTCATCCCTGCACTGATGGGAGAGATGTTC 420

Qy     421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 480
Db     421 TGGGCTGCAGCAATTTTCTCTGCATTAGGGGCTACCATCAGCGTAATCATTGATGTGGAT 480

Qy     481 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 540
Db     481 GTGAACATATCGGTCATTGTCTCCGCACTCATTGCCATTCTTTATACCCTCGTGGGAGGG 540

Qy     541 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 600
Db     541 CTCTACTCTGTGGCATATACTGATGTTGTACAGCTATTCTGCATTTTTATAGGATTGTGG 600

Qy     601 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 660
Db     601 ATCAGTGTCCCATTTGCCCTGTCACATCCTGCAGTCACCGACATTGGATTCACTGCTGTG 660

Qy     661 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 720
Db     661 CATGCTAAATACCAGAGTCCCTGGCTGGGAACCATTGAATCAGTTGAAGTCTACACCTGG 720

Qy     721 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 780
Db     721 CTTGATAATTTTCTGTTGTTGATGCTGGGTGGAATACCATGGCAAGCCTACTTCCAGAGG 780

Qy     781 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 840
Db     781 GTCCTCTCTTCATCGTCAGCGACCTATGCTCAGGTGCTGTCCTTCCTGGCAGCTTTTGGG 840

Qy     841 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 900
Db     841 TGCCTGGTGATGGCTCTACCAGCCATTTGCATTGGGGCCATTGGAGCCTCCACAGACTGG 900

Qy     901 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 960
Db     901 AACCAAACTGCATATGGGTTTCCAGATCCCAAGACCAAGGAGGAAGCAGACATGATTCTC 960

Qy     961 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1020
Db     961 CCGATTGTTCTACAGTACCTCTGCCCTGTGTACATTTCCTTCTTTGGGCTTGGTGCTGTT 1020

Qy    1021 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1080
Db    1021 TCTGCTGCTGTCATGTCCTCGGCTGACTCATCCATCCTATCAGCAAGTTCCATGTTTGCT 1080

Qy    1081 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140
Db    1081 CGGAATATCTACCAGCTTTCCTTCAGACAAAATGCATCAGACAAGGAAATTGTGTGGGTC 1140

Qy    1141 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1200
Db    1141 ATGAGGATCACTGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTCACGAAG 1200

Qy    1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1260
Db    1201 ACTGTGTATGGGCTCTGGTACCTGAGCTCTGACCTTGTCTACATCATCATCTTCCCACAG 1260

Qy    1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1320
Db    1261 CTGCTCTGTGTACTCTTCATCAAAGGAACCAACACTTATGGGGCAGTTGCTGGTTATATT 1320

Qy    1321 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1380
Db    1321 TTTGGACTTTTCCTGAGAATTACCGGAGGAGAGCCATATCTATACTTGCAGCCCTTAATC 1380

Qy    1381 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1440
Db    1381 TTCTACCCTGGTTATTACCCTGACAAGAATGGTATATACAATCAGAGGTTCCCATTTAAA 1440

Qy    1441 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1500
Db    1441 ACTCTCTCCATGGTTACCTCATTCTTTACCAACATTTGTGTTTCCTATCTAGCCAAGTAT 1500

Qy    1501 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1560
Db    1501 CTATTTGAAAGTGGAACCTTGCCTCCAAAATTAGATATATTTGATGCTGTTGTCTCAAGG 1560

Qy    1561 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1620
Db    1561 CACAGTGAAGAGAACATGGACAAGACCATTCTAGTCAGAAATGAAAACATCAAATTAAAT 1620

Qy    1621 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1680
Db    1621 GAACTTGCACCTGTAAAGCCTCGACAGAGCCTAACCCTCAGTTCAACTTTCACCAATAAA 1680

Qy    1681 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1740
Db    1681 GAGGCTCTCCTTGATGTTGATTCCAGTCCAGAGGGATCTGGGACTGAAGATAACTTACAA 1740

Qy    1741 TGA 1743
Db    1741 TGA 1743
```

RESULT 7
BD005266
ID BD005266 standard; DNA; ROD; 1743 BP.
XX
AC BD005266;
XX
SV BD005266.1
XX
DT 08-FEB-2002 (Rel. 70, Created)
DT 08-FEB-2002 (Rel. 70, Last updated, Version 1)
XX

```
DE   High-affinity choline transporter.
XX
KW   JP 03074455-T/2.
XX
OS   Rattus norvegicus (Norway rat)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
XX
RN   [1]
RP   1-1743
RA   Haga T., Okuda T.;
RT   "High-affinity choline transporter";
RL   Patent number JP03074455-T/2, 19-JAN-2001.
RL   JAPAN SCIENCE AND TECHNOLOGY CORP,TATSUYA HAGA,TAKASHI OKUDA.
XX
CC   OS   Rattus norvegicus (rat)
CC   PN   JP 03074455-T/2
CC   PD   19-JAN-2001
CC   PF   18-AUG-2000 JP 2000005545
CC   PR   27-AUG-1999 JP 99P   240642,27-DEC-1999 JP 99P   368991
CC   PI   TATSUYA HAGA,TAKASHI OKUDA
CC   PC   C12N15/12,C07K14/47,C12Q1/68,C07K19/00,C07K16/18,C12N5/10,
CC   PC   A61K38/17,
CC   PC   A61K45/00,A61P25/28,G01N33/53,A01K67/027
CC   CC
CC   FH   Key             Location/Qualifiers
CC   FT   CDS             (1). .(1743)
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .1743
FT                   /db_xref="taxon:10116"
FT                   /organism="Rattus norvegicus"
XX
SQ   Sequence 1743 BP; 414 A; 402 C; 404 G; 523 T; 0 other;

  Query Match           80.0%;  Score 1394.2;  DB 23;  Length 1743;
  Best Local Similarity 87.5%;  Pred. No. 0;
  Matches 1525;  Conservative    0;  Mismatches  218;  Indels    0;  Gaps    0;

Qy      1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60
          ||| |||||||||| ||||||||  |||| || ||| ||||||||||||| || || |||
Db      1 ATGCCTTTCCATGTAGAAGGACTAGTAGCGATTATCCTGTTCTACCTTCTTATATTTCTG 60

Qy     61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120
          ||||||||||||||||| |||| |||||||||||| || |  |||||||| |||||||||
Db     61 GTTGGAATATGGGCTGCATGGAAAACCAAAAACAGCGGTAATGCAGAAGAACGCAGCGAA 120

Qy    121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180
          ||||||||||||| |||||||| |||||||| |||||||||| ||||||||||||||||||
Db    121 GCCATCATAGTTGGGGGCCGAGACATTGGTTTGTTGGTTGGTGGTTTTACCATGACAGCC 180

Qy    181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240
          |||||||| |||||||| || ||||| || |||||||||||||||||||  ||||||| |
Db    181 ACCTGGGTTGGAGGAGGTTACATCAACGGGACAGCTGAAGCAGTTTATGGGCCAGGTTGT 240

Qy    241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300
          || |||||||||||||||||||| |||||||||||||| ||||||||||||||||||||||
Db    241 GGTCTAGCTTGGGCTCAGGCACCCATTGGATATTCTCTGAGTCTGATTTTAGGTGGCCTG 300

Qy    301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360
          || |||||||||||||||||||| ||||| |||||||| |||||||||||||||||||| ||
Db    301 TTTTTTGCAAAACCTATGCGTTCCAAGGGATATGTGACTATGTTAGACCCGTTTCAACAG 360

Qy    361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420
          |||||||||| |||||||| || || |||| || |||||||||||||||||| ||||||
Db    361 ATCTATGGAAAGCGCATGGGTGGGCTGCTGTTCATCCCTGCACTGATGGGAGAGATGTTC 420

Qy    421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 480
          |||||||||||||||||||||| || || || |||||||||| |||| |||||||||||
Db    421 TGGGCTGCAGCAATTTTCTCTGCATTAGGGGCTACCATCAGCGTAATCATTGATGTGGAT 480

Qy    481 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 540
          || |||| || |||||  |||| |||||||||||||| ||| || || || |||||||||
Db    481 GTGAACATATCGGTCATTGTCTCCGCACTCATTGCCATTCTTTATACCCTCGTGGGAGGG 540

Qy    541 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 600
          ||||| |||||||| || |||||||| || ||||| || ||||||||| ||||  |||||
Db    541 CTCTACTCTGTGGCATATACTGATGTTGTACAGCTATTCTGCATTTTTATAGGATTGTGG 600

Qy    601 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 660
          ||||| ||||| |||||  |||||||||||||||||| | ||||| || ||||||||||||
Db    601 ATCAGTGTCCCATTTGCCCTGTCACATCCTGCAGTCACCGACATTGGATTCACTGCTGTG 660

Qy    661 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 720
          ||||| |||||||| |  || |||||||||||  |||| |||  |||||||||| | |||
Db    661 CATGCTAAATACCAGAGTCCCTGGCTGGGAACCATTGAATCAGTTGAAGTCTACACCTGG 720

Qy    721 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 780
          ||||||| |||||||||||||||||||||||||||||||| |||||||||| ||||| ||||||
Db    721 CTTGATAATTTTCTGTTGTTGATGCTGGGTGGAATACCATGGCAAGCCTACTTCCAGAGG 780

Qy    781 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 840
          || |||||||| || ||||| |||||||||| |||||||||||||||||||||||||| |||
Db    781 GTCCTCTCTTCATCGTCAGCGACCTATGCTCAGGTGCTGTCCTTCCTGGCAGCTTTTGGG 840

Qy    841 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 900
          |||||||||||||  | ||||||||  |||||||||||||||||||| || |||||||||
Db    841 TGCCTGGTGATGGCTCTACCAGCCATTTGCATTGGGGCCATTGGAGCCTCCACAGACTGG 900

Qy    901 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 960
          ||||| |||||||||||| |||||||||||||||| |  || || ||||||||||||| |
Db    901 AACCAAACTGCATATGGGTTTCCAGATCCCAAGACCAAGGAGGAAGCAGACATGATTCTC 960

Qy    961 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1020
          || |||||||| |||| |||||||||||||| |||| |||||||| |||||||| |||||
Db    961 CCGATTGTTCTACAGTACCTCTGCCCTGTGTACATTTCCTTCTTTGGGCTTGGTGCTGTT 1020

Qy   1021 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1080
          |||||||||||| |||| || || || || |||||| | ||||||||||||||||||||
Db   1021 TCTGCTGCTGTCATGTCCTCGGCTGACTCATCCATCCTATCAGCAAGTTCCATGTTTGCT 1080

Qy   1081 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140
          ||||| |||||||||||||||||||||||||||||| || ||||| ||||| || ||||||
Db   1081 CGGAATATCTACCAGCTTTCCTTCAGACAAAATGCATCAGACAAGGAAATTGTGTGGGTC 1140

Qy   1141 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1200
          ||| | ||||| ||||||||||||||||||||||||||||||||||||||||||| |||||
Db   1141 ATGAGGATCACTGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTCACGAAG 1200

Qy   1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1260
          |||||||||||||||||||||| || |||||||||| |||||| | |||||||| |||
Db   1201 ACTGTGTATGGGCTCTGGTACCTGAGCTCTGACCTTGTCTACATCATCATCTTCCCACAG 1260

Qy   1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1320
          ||||| |||||||||||  | || |||||||||| |||||||| || || |||||| ||
Db   1261 CTGCTCTGTGTACTCTTCATCAAAGGAACCAACACTTATGGGGCAGTTGCTGGTTATATT 1320

Qy   1321 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1380
          | ||| || |||||||||||| || ||||| |||||||||| ||  | |||||||| |||
Db   1321 TTTGGACTTTTCCTGAGAATTACCGGAGGAGAGCCATATCTATACTTGCAGCCCTTAATC 1380

Qy   1381 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1440
          |||||||||| |||||||||  | |||||||||| |||||||  || |||||||||
Db   1381 TTCTACCCTGGTTATTACCCTGACAAGAATGGTATATACAATCAGAGGTTCCCATTTAAA 1440

Qy   1441 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1500
          || ||  |||||||||| |||||||| |||||||||||  | ||||||||||||||||||
Db   1441 ACTCTCTCCATGGTTACCTCATTCTTTACCAACATTTGTGTTTCCTATCTAGCCAAGTAT 1500

Qy   1501 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1560
          |||||||||||||||||||||| || |||||||| |||||||||||||||  ||||
Db   1501 CTATTTGAAAGTGGAACCTTGCCTCCAAAATTAGATATATTTGATGCTGTTGTCTCAAGG 1560

Qy   1561 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1620
          |||||||||||| |||||||| ||||| ||||| |||| ||||||||| || |||||| ||
```

```
Db    1561 CACAGTGAAGAGAACATGGACAAGACCATTCTAGTCAGAAATGAAAACATCAAATTAAAT 1620

Qy    1621 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1680
           |||||||||| ||| ||||| |||||||||  | ||||||||  ||||||||||||||||||
Db    1621 GAACTTGCACCTGTAAAGCCTCGACAGAGCCTAACCCTCAGTTCAACTTTCACCAATAAA 1680

Qy    1681 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1740
           |||||  ||||||||||||||||||||||||| || ||||||||||||||||| ||||||
Db    1681 GAGGCTCTCCTTGATGTTGATTCCAGTCCAGAGGGATCTGGGACTGAAGATAACTTACAA 1740

Qy    1741 TGA 1743
           |||
Db    1741 TGA 1743
```

```
RESULT 8
AB030947
LOCUS       AB030947                4904 bp    mRNA    linear   ROD 03-FEB-2000
DEFINITION  Rattus norvegicus mRNA for high-affinity choline transporter CHT1,
            complete cds.
ACCESSION   AB030947
VERSION     AB030947.1  GI:6863033
KEYWORDS    choline transporter; high-affinity choline transporter CHT1.
SOURCE      Rattus norvegicus (strain:Wistar) adult spinal cord cDNA to mRNA,
            clone_lib:rat spinal cord cDNA library clone:CHT1.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (sites)
  AUTHORS   Okuda,T., Haga,T., Kanai,Y., Endou,H., Ishihara,T. and Katsura,I.
  TITLE     Identification and characterization of the high-affinity choline
            transporter
  JOURNAL   Nat. Neurosci. 3 (2), 120-125 (2000)
  MEDLINE   20116099
REFERENCE   2  (bases 1 to 4904)
  AUTHORS   Okuda,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-1999) Takashi Okuda, University of Tokyo, Faculty
            of Medicine, Department of Neurochemistry; Hongo 7-3-1, Bunkyo-ku
            113-0033, Japan (E-mail:okuda@m.u-tokyo.ac.jp, Tel:+81-3-5841-3560,
            Fax:+81-3-3814-8154)
COMMENT     Sequence updated (11-Jan-2000).
FEATURES             Location/Qualifiers
     source          1. .4904
                     /organism="Rattus norvegicus"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /clone="CHT1"
                     /tissue_type="spinal cord"
                     /clone_lib="rat spinal cord cDNA library"
                     /dev_stage="adult"
     CDS             224. .1966
                     /codon_start=1
                     /product="high-affinity choline transporter CHT1"
                     /protein_id="BAA90484.1"
                     /db_xref="GI:6863034"
                     /translation="MPFHVEGLVAIILFYLLIFLVGIWAAWKTKNSGNAEERSEAIIV
                     GGRDIGLLVGGFTMTATWVGGGYINGTAEAVYGPGCGLAWAQAPIGYSLSLILGGLFF
                     AKPMRSKGYVTMLDPFQQIYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVD
                     VNISVIVSALIAILYTLVGGLYSVAYTDVVQLFCIFIGLWISVPFALSHPAVTDIGFT
                     AVHAKYQSPWLGTIESVEVYTWLDNFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFL
                     AAFGCLVMALPAICIGAIGASTDWNQTAYGFPDPKTKEEADMILPIVLQYLCPVYISF
                     FGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITVFVFGASA
                     TAMALLTKTVYGLWYLSSDLVYIIIFPQLLCVLFIKGTNTYGAVAGYIFGLFLRITGG
                     EPYLYLQPLIFYPGYYPDKNGIYNQRFPFKTLSMVTSFFTNICVSYLAKYLFESGTLP
                     PKLDIFDAVVSRHSEENMDKTILVRNENIKLNELAPVKPRQSLTLSSTFTNKEALLDV
                     DSSPEGSGTEDNLQ"
BASE COUNT     1447 a    991 c    939 g   1527 t
ORIGIN
```

```
Query Match             80.0%;  Score 1394.2;  DB 10;  Length 4904;
Best Local Similarity   87.5%;  Pred. No. 0;

  Matches 1525;  Conservative    0;  Mismatches  218;  Indels    0;  Gaps    0;

Qy       1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60
           ||| |||||||||| ||||||||  |||| || ||| |||||||||||||| || || |||
Db     224 ATGCCTTTCCATGTAGAAGGACTAGTAGCGATTATCCTGTTCTACCTTCTTATATTTCTG 283

Qy      61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120
           ||||||||||||||||| |||| |||||||||||| || |  |||||||| |||||||||
Db     284 GTTGGAATATGGGCTGCATGGAAAACCAAAAACAGCGGTAATGCAGAAGAACGCAGCGAA 343

Qy     121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180
           ||||||||||||||| |||||||| |||||||| |||||||||| ||||||||||||||
Db     344 GCCATCATAGTTGGGGGCCGAGACATTGGTTTGTTGGTTGGTGGTTTTACCATGACAGCC 403

Qy     181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240
           |||||||| |||||||| || ||||| || ||||||||||||||||||| ||||||| |
Db     404 ACCTGGGTTGGAGGAGGTTACATCAACGGGACAGCTGAAGCAGTTTATGGGCCAGGTTGT 463

Qy     241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300
           || |||||||||||||||||||| |||||||||||||| ||||||||||||||||||||
Db     464 GGTCTAGCTTGGGCTCAGGCACCCATTGGATATTCTCTGAGTCTGATTTTAGGTGGCCTG 523

Qy     301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360
           || |||||||||||||||||||| ||||| |||||||| ||||||||||||||||||  ||
Db     524 TTTTTTGCAAAACCTATGCGTTCCAAGGGATATGTGACTATGTTAGACCCGTTTCAACAG 583

Qy     361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420
           |||||||||| |||||||| || || ||||| || |||||||||||||||||| ||||||
Db     584 ATCTATGGAAAGCGCATGGGTGGGCTGCTGTTCATCCCTGCACTGATGGGAGAGATGTTC 643

Qy     421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 480
           |||||||||||||||||||||||||| || || || |||||||||| ||||| |||||||||
Db     644 TGGGCTGCAGCAATTTTCTCTGCATTAGGGGCTACCATCAGCGTAATCATTGATGTGGAT 703

Qy     481 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 540
           || |||| || |||||  |||| |||||||||||| ||| || || || |||||||||
Db     704 GTGAACATATCGGTCATTGTCTCCGCACTCATTGCCATTCTTTATACCCTCGTGGGAGGG 763

Qy     541 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 600
           ||||| |||||||| || |||||||| || ||||| || ||||||||| ||||  |||||
Db     764 CTCTACTCTGTGGCATATACTGATGTTGTACAGCTATTCTGCATTTTTATAGGATTGTGG 823

Qy     601 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 660
           ||||| ||||| |||||  ||||||||||||||||| | ||||| || ||||||||||||
Db     824 ATCAGTGTCCCATTTGCCCTGTCACATCCTGCAGTCACCGACATTGGATTCACTGCTGTG 883

Qy     661 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 720
           ||||| |||||||| |  || |||||||||||  |||| |||  |||||||||| | |||
Db     884 CATGCTAAATACCAGAGTCCCTGGCTGGGAACCATTGAATCAGTTGAAGTCTACACCTGG 943

Qy     721 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 780
           ||||||| ||||||||||||||||||||||||||||||||||| |||||||||| ||||||
Db     944 CTTGATAATTTTCTGTTGTTGATGCTGGGTGGAATACCATGGCAAGCCTACTTCCAGAGG 1003

Qy     781 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 840
           || |||||||| || ||||| |||||||||| |||||||||||||||||||||||||| |||
Db    1004 GTCCTCTCTTCATCGTCAGCGACCTATGCTCAGGTGCTGTCCTTCCTGGCAGCTTTTGGG 1063

Qy     841 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 900
           |||||||||||||  | ||||||||  |||||||||||||||||| || |||||||||
Db    1064 TGCCTGGTGATGGCTCTACCAGCCATTTGCATTGGGGCCATTGGAGCCTCCACAGACTGG 1123

Qy     901 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 960
           ||||| |||||||||||| |||||||||||||||| |  || || |||||||||||| |
Db    1124 AACCAAACTGCATATGGGTTTCCAGATCCCAAGACCAAGGAGGAAGCAGACATGATTCTC 1183

Qy     961 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1020
           || |||||||| ||||| |||||||||||||| ||||| |||||||| |||||||| |||
Db    1184 CCGATTGTTCTACAGTACCTCTGCCCTGTGTACATTTCCTTCTTTGGGCTTGGTGCTGTT 1243

Qy    1021 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1080
           |||||||||| ||||| || || || || |||||| | ||||||||||||||||||||
Db    1244 TCTGCTGCTGTCATGTCCTCGGCTGACTCATCCATCCTATCAGCAAGTTCCATGTTTGCT 1303
```

```
Qy   1081 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140
          ||||| |||||||||||||||||||||||||| || ||||| ||||| || ||||||
Db   1304 CGGAATATCTACCAGCTTTCCTTCAGACAAAATGCATCAGACAAGGAAATTGTGTGGGTC 1363

Qy   1141 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1200
          ||| | ||||| |||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   1364 ATGAGGATCACTGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTCACGAAG 1423

Qy   1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1260
          |||||||||||||||||||||| || |||||||||| |||||| | |||||||| |||
Db   1424 ACTGTGTATGGGCTCTGGTACCTGAGCTCTGACCTTGTCTACATCATCATCTTCCCACAG 1483

Qy   1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1320
          ||||| |||||||||||  | || |||||||||| ||||||| || || |||||| ||
Db   1484 CTGCTCTGTGTACTCTTCATCAAAGGAACCAACACTTATGGGGCAGTTGCTGGTTATATT 1543

Qy   1321 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1380
          | ||| || |||||||||| || ||||| |||||||||| ||  | |||||||| |||
Db   1544 TTTGGACTTTTCCTGAGAATTACCGGAGGAGAGCCATATCTATACTTGCAGCCCTTAATC 1603

Qy   1381 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1440
          |||||||||| |||||||||| | |||||||||| ||||||| || |||||||||
Db   1604 TTCTACCCTGGTTATTACCCTGACAAGAATGGTATATACAATCAGAGGTTCCCATTTAAA 1663

Qy   1441 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1500
          || ||  ||||||||| |||||||| |||||||||| | ||||||||||||||||||
Db   1664 ACTCTCTCCATGGTTACCTCATTCTTTACCAACATTTGTGTTTCCTATCTAGCCAAGTAT 1723

Qy   1501 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1560
          |||||||||||||||||||||| || ||||||||| |||||||||||||||  ||||
Db   1724 CTATTTGAAAGTGGAACCTTGCCTCCAAAATTAGATATATTTGATGCTGTTGTCTCAAGG 1783

Qy   1561 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1620
          |||||||||| |||||||| ||||| ||||| |||| ||||||||| || |||||| ||
Db   1784 CACAGTGAAGAGAACATGGACAAGACCATTCTAGTCAGAAATGAAAACATCAAATTAAAT 1843

Qy   1621 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1680
          |||||||||| ||| ||||| |||||||| | |||||||| ||||||||||||||||||
Db   1844 GAACTTGCACCTGTAAAGCCTCGACAGAGCCTAACCCTCAGTTCAACTTTCACCAATAAA 1903

Qy   1681 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1740
          |||||  |||||||||||||||||||||||| || |||||||||||||||| |||||
Db   1904 GAGGCTCTCCTTGATGTTGATTCCAGTCCAGAGGGATCTGGGACTGAAGATAACTTACAA 1963

Qy   1741 TGA 1743
          |||
Db   1964 TGA 1966
```

RESULT 9
AF276872

```
LOCUS       AF276872                1743 bp    mRNA    linear   ROD 28-FEB-2001
DEFINITION  Mus musculus sodium and chloride-dependent high-affinity choline
            transporter mRNA, complete cds.
ACCESSION   AF276872
VERSION     AF276872.2  GI:13162669
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1743)
  AUTHORS   Apparsundaram,S., Ferguson,S.M. and Blakely,R.D.
  TITLE     Molecular cloning and characterization of human and murine
            high-affinity choline transporters
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1743)
  AUTHORS   Apparsundaram,S., Ferguson,S.M. and Blakely,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-2000) Department of Pharmacology and Center for
            Molecular Neuroscience, Vanderbilt University, 23rd Avenue South at
            Pierce, Nashville, TN 37232-6420, USA
REFERENCE   3  (bases 1 to 1743)
  AUTHORS   Apparsundaram,S., Ferguson,S.M. and Blakely,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2001) Department of Pharmacology and Center for
            Molecular Neuroscience, Vanderbilt University, 23rd Avenue South at
            Pierce, Nashville, TN 37232-6420, USA
  REMARK    Sequence update by submitter
COMMENT     On Feb 28, 2001 this sequence version replaced gi:11527247.
FEATURES             Location/Qualifiers
     source          1. .1743
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     CDS             1. .1743
                     /codon_start=1
                     /product="sodium and chloride-dependent high-affinity
                     choline transporter"
                     /protein_id="AAG36945.2"
                     /db_xref="GI:13162670"
                     /translation="MPFHVEGLVAIILFYLLIFLVGIWAAWKTKNSGNPEERSEAIIV
                     GGRDIGLLVGGFTMTATWVGGGYINGTAEAVYGPGCGLAWAHAPIGYSLSLILGGLFF
                     AKPMRSKGYVTMLDPFKQIYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVD
                     VNISVIVSALIAILYTLVGGLYSVAYTDVVQLFCIFIGLWISVPFALSHPAVTDIGFT
                     AVHAKYQSPWLGTIESVEVYTWLDNFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFL
                     AAFGCLVMALPAICIGAIGASTDWNQTAYGYPDPKTKEEADMILPIVLQYLCPVYISF
                     FGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITVLVFGASA
                     TAMALLTKTVYGLWYLSSDLVYIIIFPQLLCVLFIKGTNTYGAVAGYIFGLFLRITGG
                     EPYLYLQPLIFYPGYYSDKNGIYNQRFPFKTLSMVTSFFTNICVSYLAKYLFESGTLP
                     PKLDVFDAVVARHSEENMDKTILVRNENIKLNELAPVKPRQSLTLSSTFTNKEALLDV
                     DSSPEGSGTEDNLQ"
BASE COUNT      406 a    409 c    410 g    518 t
ORIGIN
```

Query Match 78.9%; Score 1375; DB 10; Length 1743;
Best Local Similarity 86.8%; Pred. No. 0;
Matches 1513; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

```
Qy       1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60
           ||| |||||||||||||||||| ||||||| ||| | |||||||| || || || |||
Db       1 ATGCCTTTCCATGTGGAAGGACTGGTAGCTATTATCCTCTTCTACCTCCTTATATTTCTG 60

Qy      61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120
           |||||||||||||||| |||| |||||||||||| |||| | ||||||||||||| |||
Db      61 GTTGGAATATGGGCTGCATGGAAAACCAAAAACAGCGGCAACCCAGAAGAGCGCAGTGAA 120

Qy     121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180
           |||||||||| || ||||| || |||||||| |||||||||| ||||||||||||||||
Db     121 GCCATCATAGTCGGGGGCCGTGACATTGGTTTGTTGGTTGGTGGTTTTACCATGACAGCC 180

Qy     181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240
           |||||||| |||||||| || |||||||| |||| |||||||| |||| |||||||
Db     181 ACCTGGGTTGGAGGAGGCTACATCAATGGGACAGCAGAAGCAGTGTATGGGCCAGGTTGT 240

Qy     241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300
           || |||||||||||||| |||| |||||||||||||| |||| ||||||||||| |||
Db     241 GGTCTAGCTTGGGCTCATGCACCCATTGGATATTCTCTGAGTCTAATTTTAGGTGGTCTG 300

Qy     301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360
           || |||| |||||||||||||| |||| |||||||| |||||||||| ||  | ||
Db     301 TTTTTTGCGAAACCTATGCGTTCCAAGGGATATGTGACTATGTTAGACCCATTCAAACAG 360

Qy     361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420
           |||||||||| |||||||| || || || || || |||||||||||||||| ||||||
Db     361 ATCTATGGAAAGCGCATGGGTGGGCTGCTCTTCATCCCTGCACTGATGGGAGAGATGTTC 420

Qy     421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 480
           |||||||||||||||||||||||| || || |||||||||||||||||| |||||||||
Db     421 TGGGCTGCAGCAATTTTCTCTGCATTAGGGGCCACCATCAGCGTGATCATTGATGTGGAT 480

Qy     481 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 540
           || |||| || |||||  |||||||||||||||||| ||| || || || ||||| |||
Db     481 GTGAACATATCGGTCATTGTCTCTGCACTCATTGCCATTCTTTATACCCTAGTGGGTGGG 540

Qy     541 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 600
```

```
Db    541 CTCTACTCTGTGGCATATACTGATGTTGTCCAGCTATTCTGCATTTTTATAGGACTGTGG 600

Qy    601 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 660
Db    601 ATCAGTGTCCCTTTTGCCCTGTCACATCCTGCAGTCACCGACATCGGATTCACAGCTGTG 660

Qy    661 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 720
Db    661 CATGCTAAATACCAGAGTCCCTGGCTGGGAACCATTGAATCAGTTGAAGTCTACACCTGG 720

Qy    721 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 780
Db    721 CTTGATAATTTTCTGTTATTGATGCTGGGTGGAATCCCATGGCAAGCCTACTTCCAGAGG 780

Qy    781 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 840
Db    781 GTCCTCTCTTCATCCTCAGCCACCTATGCTCAGGTACTGTCCTTCCTGGCAGCTTTTGGG 840

Qy    841 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 900
Db    841 TGCCTGGTGATGGCTCTACCCGCCATATGCATAGGAGCTATTGGAGCTTCCACAGACTGG 900

Qy    901 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 960
Db    901 AACCAGACTGCCTACGGGTATCCAGATCCCAAGACTAAGGAGGAAGCAGACATGATTCTC 960

Qy    961 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1020
Db    961 CCGATCGTTCTGCAGTACCTCTGCCCTGTGTACATCTCCTTCTTTGGGCTTGGTGCTGTT 1020

Qy   1021 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1080
Db   1021 TCAGCTGCTGTCATGTCCTCAGCTGACTCGTCCATCCTGTCGGCGAGTTCTATGTTTGCT 1080

Qy   1081 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140
Db   1081 CGGAATATCTACCAGCTTTCCTTCAGACAAAATGCATCAGACAAGGAAATTGTGTGGGTC 1140

Qy   1141 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1200
Db   1141 ATGAGGATCACTGTGCTTGTGTTCGGAGCATCTGCAACAGCCATGGCTTTGCTGACGAAG 1200

Qy   1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1260
Db   1201 ACTGTGTATGGGCTCTGGTACCTGAGCTCTGACCTTGTCTACATCATCATCTTCCCACAG 1260

Qy   1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1320
Db   1261 CTGCTCTGTGTACTCTTCATCAAAGGAACCAACACTTATGGGGCAGTTGCTGGTTATATT 1320

Qy   1321 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1380
Db   1321 TTTGGACTATTCCTGAGAATTACTGGAGGAGAGCCATATCTATACTTGCAGCCCTTAATC 1380

Qy   1381 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1440
Db   1381 TTCTACCCTGGTTATTACTCTGACAAGAATGGTATATACAATCAGAGGTTCCCATTTAAA 1440

Qy   1441 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1500
Db   1441 ACTCTCTCCATGGTTACCTCATTCTTTACCAACATTTGTGTTTCTTATCTAGCCAAGTAT 1500

Qy   1501 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1560
Db   1501 CTATTTGAAAGTGGAACCTTGCCTCCAAAATTAGATGTATTTGATGCTGTTGTCGCAAGG 1560

Qy   1561 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1620
Db   1561 CACAGTGAAGAGAACATGGACAAGACCATTCTAGTCAGAAATGAAAATATCAAATTAAAT 1620

Qy   1621 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1680
```

```
Db   1621 GAACTTGCACCTGTGAAACCTCGGCAGAGCCTAACCCTCAGTTCAACTTTCACCAATAAG 1680

Qy   1681 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1740
Db   1681 GAGGCCCTCCTTGATGTTGATTCCAGTCCGGAGGGGTCTGGGACTGAAGATAATTTACAA 1740

Qy   1741 TGA 1743
Db   1741 TGA 1743


RESULT 10
BD012720
LOCUS       BD012720                1743 bp    DNA     linear   PAT 02-AUG-2002
DEFINITION  High-affinity choline transporter.
ACCESSION   BD012720
VERSION     BD012720.1  GI:22092909
KEYWORDS    WO 0116315-A/4.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1743)
  AUTHORS   Haga,T. and Okuda,T.
  TITLE     High-affinity choline transporter
  JOURNAL   Patent: WO 0116315-A 4 08-MAR-2001;
            JAPAN SCIENCE AND TECHNOLOGY CORP,TATSUYA HAGA,TAKASHI OKUDA
COMMENT     OS   Mus musculus (mouse)
            PN   WO 0116315-A/4
            PD   08-MAR-2001
            PF   18-AUG-2000 WO 2000JP005545
            PR   27-AUG-1999 JP 99P   240642,27-DEC-1999 JP 99P   368991 PI
            TATSUYA HAGA,TAKASHI OKUDA
            PC   C12N15/12,C07K14/47,C12Q1/68,C07K19/00,C07K16/18,C12N5/10, PC
             A61K38/17,
            PC   A61K45/00,A61P25/28,G01N33/53,A01K67/027
            CC
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(1743).
FEATURES            Location/Qualifiers
     source         1. .1743
                    /organism="Mus musculus"
                    /db_xref="taxon:10090"
BASE COUNT      407 a    410 c    409 g    517 t
ORIGIN

  Query Match           78.8%;  Score 1373.4;  DB 6;  Length 1743;
  Best Local Similarity  86.7%;  Pred. No. 0;
  Matches 1512;  Conservative    0;  Mismatches  231;  Indels    0;  Gaps    0;

Qy       1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60
Db       1 ATGTCTTTCCACGTAGAAGGACTGGTAGCTATTATCCTCTTCTACCTCCTTATATTTCTG 60

Qy      61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120
Db      61 GTTGGAATATGGGCTGCATGGAAAACCAAAAACAGCGGCAACCCAGAAGAGCACAGTGAA 120

Qy     121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180
Db     121 GCCATCATAGTCGGGGGCCGTGACATTGGTTTGTTGGTTGGTGGTTTTACCATGACAGCC 180

Qy     181 ACCTGGGTCGGAGGAGGGTACATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240
Db     181 ACCTGGGTTGGAGGAGGCTACATCAATGGGACAGCAGAAGCAGTGTATGGGCCAGGTTGT 240

Qy     241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300
Db     241 GGTCTAGCTTGGGCTCAGGCACCCATTGGATATTCTCTGAGTCTAATTTTAGGTGGTCTG 300

Qy     301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360
Db     301 TTTTTTGCGAAACCTATGCGTTCCAAGGGATATGTGACTATGTTAGACCCATTTCAACAG 360
```

```
Qy    361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420
Db    361 ATCTATGGAAAGCGCATGGGTGGGCTGCTCTTCATCCCTGCACTGATGGGAGAGATGTTC 420

Qy    421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 480
Db    421 TGGGCTGCAGCAATTTTCTCTGCATTAGGGGCCACCATCAGCGTGATCATTGATGTGGAT 480

Qy    481 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 540
Db    481 GTGAACATATCGGTCATTGTCTCTGCACTCATTGCCATTCTTTATACCCTAGTGGGTGGG 540

Qy    541 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 600
Db    541 CTCTACTCTGTGGCATATACTGATGTTGTCCAGCTATTCTGCATTTTTATAGGACTGTGG 600

Qy    601 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 660
Db    601 ATCAGTGTCCCTTTTGCCCTGTCACATCCTGCAGTCACCGACATCGGATTCACAGCTGTG 660

Qy    661 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 720
Db    661 CATGCTAAATACCAGAGTCCCTGGCTGGGAACCATTGAATCAGTTGAAGTCTACACCTGG 720

Qy    721 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 780
Db    721 CTTGATAATTTTCTGTTATTGATGCTGGGTGGAATCCCATGGCAAGCCTACTTCCAGAGG 780

Qy    781 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 840
Db    781 GTCCTCTCTTCATCCTCAGCCACCTATGCTCAGGTACTGTCCTTCCTGGCAGCTTTTGGG 840

Qy    841 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 900
Db    841 TGCCTGGTGATGGCTCTACCCGCCATATGCATAGGAGCTATTGGAGCTTCCACAGACTGG 900

Qy    901 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 960
Db    901 AACCAGACTGCCTACGGGTATCCAGATCCCAAGACTAAGGAGGAAGCAGACATGATTCTC 960

Qy    961 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1020
Db    961 CCGATCGTTCTGCAGTACCTCTGCCCTGTGTACATCTCCTTCTTTGGGCTTGGTGCTGTT 1020

Qy   1021 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1080
Db   1021 TCAGCTGCTGTCATGTCCTCAGCTGACTCGTCCATCCTGTCGGCGAGTTCTATGTTTGCT 1080

Qy   1081 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140
Db   1081 CGGAATATCTACCAGCTTTCCTTCAGACAAAATGCATCAGACAAGGAAATTGTGTGGGTC 1140

Qy   1141 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1200
Db   1141 ATGAGGATCACTGTGCTTGTGTTCGGAGCATCTGCAACAGCCATGGCTTTGCTGACGAAG 1200

Qy   1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1260
Db   1201 ACTGTGTATGGGCTCTGGTACCTGAGCTCTGACCTTGTCTACATCATCATCTTCCCACAG 1260

Qy   1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1320
Db   1261 CTGCTCTGTGTACTCTTCATCAAAGGAACCAACACTTATGGGGCAGTTGCTGGTTATATT 1320

Qy   1321 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1380
Db   1321 TTTGGACTATTCCTGAGAATTACTGGAGGAGAGCCATATCTATACTTGCAGCCCTTAATC 1380

Qy   1381 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1440
Db   1381 TTCTACCCTGGTTATTACTCTGACAAGAATGGTATATACAATCAGAGGTTCCCATTTAAA 1440

Qy   1441 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1500
Db   1441 ACTCTCTCCATGGTTACCTCATTCTTTACCAACATTTGTGTTTCTTATCTAGCCAAGTAT 1500

Qy   1501 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1560
Db   1501 CTATTTGAAAGTGGAACCTTGCCTCCAAAATTAGATGTATTTGATGCTGTTGTCGCAAGG 1560

Qy   1561 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1620
Db   1561 CACAGTGAAGAGAACATGGACAAGACCATTCTAGTCAGAAATGAAAATATCAAATTAAAT 1620

Qy   1621 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1680
Db   1621 GAACTTGCACCTGTGAAACCTCGGCAGAGCCTAACCCTCAGTTCAACTTTCACCAATAAG 1680

Qy   1681 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1740
Db   1681 GAGGCCCTCCTTGATGTTGATTCCAGTCCGGAGGGGTCTGGGACTGAAGATAACTTACAA 1740

Qy   1741 TGA 1743
Db   1741 TGA 1743
```

RESULT 11
BD005268
ID BD005268 standard; DNA; MUS; 1743 BP.
XX
AC BD005268;
XX
SV BD005268.1
XX
DT 08-FEB-2002 (Rel. 70, Created)
DT 08-FEB-2002 (Rel. 70, Last updated, Version 1)
XX
DE High-affinity choline transporter.
XX
KW JP 03074455-T/4.
XX
OS Mus musculus (house mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
XX
RN [1]
RP 1-1743
RA Haga T., Okuda T.;
RT "High-affinity choline transporter";
RL Patent number JP03074455-T/4, 19-JAN-2001.
RL JAPAN SCIENCE AND TECHNOLOGY CORP,TATSUYA HAGA,TAKASHI OKUDA.
XX
CC OS Mus musculus (mouse)
CC PN JP 03074455-T/4
CC PD 19-JAN-2001
CC PF 18-AUG-2000 JP 2000005545
CC PR 27-AUG-1999 JP 99P 240642,27-DEC-1999 JP 99P 368991
CC PI TATSUYA HAGA,TAKASHI OKUDA
CC PC C12N15/12,C07K14/47,C12Q1/68,C07K19/00,C07K16/18,C12N5/10,
CC PC A61K38/17,
CC PC A61K45/00,A61P25/28,G01N33/53,A01K67/027
CC CC
CC FH Key Location/Qualifiers
CC FT CDS (1). .(1743)
XX
FH Key Location/Qualifiers
FH
FT source 1. .1743
FT /db_xref="taxon:10090"
FT /organism="Mus musculus"
XX
SQ Sequence 1743 BP; 407 A; 410 C; 409 G; 517 T; 0 other;

Query Match 78.8%; Score 1373.4; DB 23; Length 1743;

Best Local Similarity 86.7%; Pred. No. 0;
Matches 1512; Conservative 0; Mismatches 231; Indels 0; Gaps 0;

```
Qy       1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60
Db       1 ATGTCTTTCCACGTAGAAGGACTGGTAGCTATTATCCTCTTCTACCTCCTTATATTTCTG 60

Qy      61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120
Db      61 GTTGGAATATGGGCTGCATGGAAAACCAAAAACAGCGGCAACCCAGAAGAGCACAGTGAA 120

Qy     121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180
Db     121 GCCATCATAGTCGGGGGCCGTGACATTGGTTTGTTGGTTGGTGGTTTTACCATGACAGCC 180

Qy     181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240
Db     181 ACCTGGGTTGGAGGAGGCTACATCAATGGGACAGCAGAAGCAGTGTATGGGCCAGGTTGT 240

Qy     241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300
Db     241 GGTCTAGCTTGGGCTCAGGCACCCATTGGATATTCTCTGAGTCTAATTTTAGGTGGTCTG 300

Qy     301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360
Db     301 TTTTTTGCGAAACCTATGCGTTCCAAGGGATATGTGACTATGTTAGACCCATTTCAACAG 360

Qy     361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420
Db     361 ATCTATGGAAAGCGCATGGGTGGGCTGCTCTTCATCCCTGCACTGATGGGAGAGATGTTC 420

Qy     421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 480
Db     421 TGGGCTGCAGCAATTTTCTCTGCATTAGGGGCCACCATCAGCGTGATCATTGATGTGGAT 480

Qy     481 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 540
Db     481 GTGAACATATCGGTCATTGTCTCTGCACTCATTGCCATTCTTTATACCCTAGTGGGTGGG 540

Qy     541 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 600
Db     541 CTCTACTCTGTGGCATATACTGATGTTGTCCAGCTATTCTGCATTTTTATAGGACTGTGG 600

Qy     601 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 660
Db     601 ATCAGTGTCCCTTTTGCCCTGTCACATCCTGCAGTCACCGACATCGGATTCACAGCTGTG 660

Qy     661 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 720
Db     661 CATGCTAAATACCAGAGTCCCTGGCTGGGAACCATTGAATCAGTTGAAGTCTACACCTGG 720

Qy     721 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 780
Db     721 CTTGATAATTTTCTGTTATTGATGCTGGGTGGAATCCCATGGCAAGCCTACTTCCAGAGG 780

Qy     781 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 840
Db     781 GTCCTCTCTTCATCCTCAGCCACCTATGCTCAGGTACTGTCCTTCCTGGCAGCTTTTGGG 840

Qy     841 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 900
Db     841 TGCCTGGTGATGGCTCTACCCGCCATATGCATAGGAGCTATTGGAGCTTCCACAGACTGG 900

Qy     901 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 960
Db     901 AACCAGACTGCCTACGGGTATCCAGATCCCAAGACTAAGGAGGAAGCAGACATGATTCTC 960

Qy     961 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1020
Db     961 CCGATCGTTCTGCAGTACCTCTGCCCTGTGTACATCTCCTTCTTTGGGCTTGGTGCTGTT 1020

Qy    1021 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1080
```

```
Db    1021 TCAGCTGCTGTCATGTCCTCAGCTGACTCGTCCATCCTGTCGGCGAGTTCTATGTTTGCT 1080

Qy    1081 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140
Db    1081 CGGAATATCTACCAGCTTTCCTTCAGACAAAATGCATCAGACAAGGAAATTGTGTGGGTC 1140

Qy    1141 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1200
Db    1141 ATGAGGATCACTGTGCTTGTGTTCGGAGCATCTGCAACAGCCATGGCTTTGCTGACGAAG 1200

Qy    1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1260
Db    1201 ACTGTGTATGGGCTCTGGTACCTGAGCTCTGACCTTGTCTACATCATCATCTTCCCACAG 1260

Qy    1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1320
Db    1261 CTGCTCTGTGTACTCTTCATCAAAGGAACCAACACTTATGGGGCAGTTGCTGGTTATATT 1320

Qy    1321 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1380
Db    1321 TTTGGACTATTCCTGAGAATTACTGGAGGAGAGCCATATCTATACTTGCAGCCCTTAATC 1380

Qy    1381 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1440
Db    1381 TTCTACCCTGGTTATTACTCTGACAAGAATGGTATATACAATCAGAGGTTCCCATTTAAA 1440

Qy    1441 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1500
Db    1441 ACTCTCTCCATGGTTACCTCATTCTTTACCAACATTTGTGTTTCTTATCTAGCCAAGTAT 1500

Qy    1501 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1560
Db    1501 CTATTTGAAAGTGGAACCTTGCCTCCAAAATTAGATGTATTTGATGCTGTTGTCGCAAGG 1560

Qy    1561 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1620
Db    1561 CACAGTGAAGAGAACATGGACAAGACCATTCTAGTCAGAAATGAAAATATCAAATTAAAT 1620

Qy    1621 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1680
Db    1621 GAACTTGCACCTGTGAAACCTCGGCAGAGCCTAACCCTCAGTTCAACTTTCACCAATAAG 1680

Qy    1681 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1740
Db    1681 GAGGCCCTCCTTGATGTTGATTCCAGTCCGGAGGGGTCTGGGACTGAAGATAACTTACAA 1740

Qy    1741 TGA 1743
Db    1741 TGA 1743
```

RESULT 12
AX080443
LOCUS AX080443 4938 bp DNA linear PAT 22-FEB-2001
DEFINITION Sequence 1 from Patent WO0078950.
ACCESSION AX080443
VERSION AX080443.1 GI:13159872
KEYWORDS .
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 4938)
AUTHORS Sierzega,M. and Albrandt,K.
TITLE Differentially expressed genes in the adipocytes of obese mice
JOURNAL Patent: WO 0078950-A 1 28-DEC-2000;
AMYLIN PHARMACEUTICALS, INC. (US)
FEATURES Location/Qualifiers
source 1. .4938
/organism="Mus musculus"
/db_xref="taxon:10090"
/note="P4P6B1"
BASE COUNT 1436 a 1012 c 976 g 1514 t

ORIGIN

Query Match 78.8%; Score 1373.4; DB 6; Length 4938;
Best Local Similarity 86.7%; Pred. No. 0;
Matches 1512; Conservative 0; Mismatches 231; Indels 0; Gaps 0;

```
Qy      1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60
Db    247 ATGTCTTTCCACGTAGAAGGACTGGTAGCTATTATCCTCTTCTACCTCCTTATACTTCTG 306

Qy     61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120
Db    307 GTTGGAATATGGGCTGCATGGAAAACCAAAAACAGCGGCAACCCAGAAGAGCGCAGTGAA 366

Qy    121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180
Db    367 GCCATCATAGTCGGGGGCCGTGACATTGGTTTGTTGGTTGGTGGTTTTACCATGACAGCC 426

Qy    181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240
Db    427 ACCTGGGTTGGAGGAGGCTACATCAATGGGACAGCAGAAGCAGTGTATGGGCCAGGTTGT 486

Qy    241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300
Db    487 GGTCTAGCTTGGGCTCAGGCACCCATTGGATATTCTCTGAGTCTAATTTTAGGTGGTCTG 546

Qy    301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360
Db    547 TTTTTTGCGAAACCTATGCGTTCCAAGGGATATGTGACTATGTTAGACCCATTTCAACAG 606

Qy    361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420
Db    607 ATCTATGGAAAGCGCATGGGTGGGCTGCTCTTCATCCCTGCACTGATGGGAGAGATGTTC 666

Qy    421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 480
Db    667 TGGGCTGCAGCAATTTTCTCTGCATTAGGGGCCACCATCAGCGTGATCATTGATGTGGAT 726

Qy    481 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 540
Db    727 GTGAACATATCGGTCATTGTCTCTGCACTCATTGCCATTCTTTATACCCTAGTGGGTGGG 786

Qy    541 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 600
Db    787 CTCTACTCTGTGGCATATACTGATGTTGTCCAGCTATTCTGCATTTTTATAGGACTGTGG 846

Qy    601 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 660
Db    847 ATCAGTGTCCCTTTTGCCCTGTCACATCCTGCAGTCACCGACATCGGATTCACAGCTGTG 906

Qy    661 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 720
Db    907 CATGCTAAATACCAGAGTCCCTGGCTGGGAACCATTGAATCAGTTGAAGTCTACACCTGG 966

Qy    721 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 780
Db    967 CTTGATAATTTTCTGTTATTGATGCTGGGTGGAATCCCATGGCAAGCCTACTTCCAGAGG 1026

Qy    781 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 840
Db   1027 GTCCTCTCTTCATCCTCAGCCACCTATGCTCAGGTACTGTCCTTCCTGGCAGCTTTTGGG 1086

Qy    841 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 900
Db   1087 TGCCTGGTGATGGCTCTACCCGCCATATGCATAGGAGCTATTGGAGCTTCCACAGACTGG 1146

Qy    901 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 960
Db   1147 AACCAGACTGCCTACGGGTATCCAGATCCCAAGACTAAGGAGGAAGCAGACATGATTCTC 1206

Qy    961 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1020
Db   1207 CCGATCGTTCTGCAGTACCTCTGCCCTGTGTACATCTCCTTCTTTGGGCTTGGTGCTGTT 1266

Qy   1021 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1080
Db   1267 TCAGCTGCTGTCATGTCCTCAGCTGACTCGTCCATCCTGTCGGCGAGTTCTATGTTTGCT 1326

Qy   1081 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140
Db   1327 CGGAATATCTACCAGCTTTCCTTCAGACAAAATGCATCAGACAAGGAAATTGTGTGGGTC 1386

Qy   1141 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1200
Db   1387 ATGAGGATCACTGTGCTTGTGTTCGGAGCATCTGCAACAGCCATGGCTTTGCTGACGAAG 1446

Qy   1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1260
Db   1447 ACTGTGTATGGGCTCTGGTACCTGAGCTCTGACCTTGTCTACATCATCATCTTCCCACAG 1506

Qy   1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1320
Db   1507 CTGCTCTGTGTACTCTTCATCAAAGGAACCAACACTTATGGGGCAGTTGCTGGTTATATT 1566

Qy   1321 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1380
Db   1567 TTTGGACTATTCCTGAGAATTACTGGAGGAGAGCCATATCTATACTTGCAGCCCTTAATC 1626

Qy   1381 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1440
Db   1627 TTCTACCCTGGTTATTACTCTGACAAGAATGGTATATACAATCAGAGGTTCCCATTTAAA 1686

Qy   1441 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1500
Db   1687 ACTCTCTCCATGGTTACCTCATTCTTTACCAACATTTGTGTTTCTTATCTAGCCAAGTAT 1746

Qy   1501 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1560
Db   1747 CTATTTGAAAGTGGAACCTTGCCTCCAAAATTAGATGTATTTGATGCTGTTGTCGCAAGG 1806

Qy   1561 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1620
Db   1807 CACAGTGAAGAGAACATGGACAAGACCATTCTAGTCAGAAATGAAAATATCAAATTAAAT 1866

Qy   1621 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1680
Db   1867 GAACTTGCACCTGTGAAACCTCGGCAGAGCCTAACCCTCAGTTCAACTTTCACCAATAAG 1926

Qy   1681 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1740
Db   1927 GAGGCCCTCCTTGATGTTGATTCCAGTCCGGAGGGGTCTGGGACTGAAGATAACTTACAA 1986

Qy   1741 TGA 1743
Db   1987 TGA 1989
```

RESULT 13
MMU401467
LOCUS MMU401467 1743 bp mRNA linear ROD 16-AUG-2000
DEFINITION Mus musculus mRNA for high affinity choline transporter (CHT1 gene).
ACCESSION AJ401467
VERSION AJ401467.1 GI:9843808
KEYWORDS ChT1 gene; high affinity choline transporter.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1743)
AUTHORS Wieland,A., Bonisch,H. and Bruss,M.
TITLE Molecular cloning of the human and murine high affinity choline transporters and characterization ofthe human gene-structure
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1743)
AUTHORS Bruess,M.

```
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-2000) Bruess M., University of Bonn, Pharmacology
            and Toxicology, Reuter str. 2b, D-53113 Bonn, GERMANY
FEATURES             Location/Qualifiers
     source          1. .1743
                     /organism="Mus musculus"
                     /strain="BALB/cJ"
                     /db_xref="taxon:10090"
                     /tissue_type="brainstem"
     gene            1. .1743
                     /gene="CHT1"
     CDS             1. .1743
                     /gene="CHT1"
                     /function="sodium- and chloride-dependent reuptake of
                     choline"
                     /codon_start=1
                     /evidence=experimental
                     /product="high affinity choline transporter"
                     /protein_id="CAC03719.1"
                     /db_xref="GI:9843809"
                     /translation="MSFHVEGLVAIILFYLLIFLVGIWAAWKTKNSGNPEEHSEAIIV
                     GGRDIGLLVGGFTMTATWVGGGYINGTAVAVYGPGCGLAWAQAPIGYSLSLILGGLFF
                     AKPMRSKGYVTMLDPFQQIYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVD
                     VNISVIVSALIAILYTLVGGLYSVAYTDVVQLFCIFIGLWISVPFALSHPAVTDIGFT
                     AVHAKYQSPWLGTIESVEVYTWLDNFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSYL
                     AAFGCLVMALPAICIGAIGASTDWNQTAYGYPDPKTKEEADMILPIVLQYLCPVYISF
                     FGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITVLVFGASA
                     TAMALLTKTVYGLWYLSSDLVYIIIFPQLLCVLFIKGTNTYGAVAGYIFGLFLRITGG
                     EPYLYLQPLIFYPGYYSDKNGIYNQRFPFKTLSMVTSFFTNICVSYLAKYLFESGTLP
                     PKLDVFDAVVARHSEENMDKTILVRNENIKLNELAPVKPRQSLTLSSTFTNKEALLDV
                     DSSPEGSGTEDNLQ"
BASE COUNT      407 a    410 c    410 g    516 t
ORIGIN

  Query Match           78.4%;  Score 1367;  DB 10;  Length 1743;
  Best Local Similarity  86.5%;  Pred. No. 0;
  Matches 1508;  Conservative    0;  Mismatches  235;  Indels    0;  Gaps    0;

Qy       1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60
Db       1 ATGTCTTTCCACGTAGAAGGACTGGTAGCTATTATCCTCTTCTACCTCCTTATATTTCTG 60

Qy      61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCAGCGCAGAAGAGCGCAGCGAA 120
Db      61 GTTGGAATATGGGCTGCATGGAAAACCAAAAACAGCGGCAACCCAGAAGAGCACAGTGAA 120

Qy     121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180
Db     121 GCCATCATAGTCGGGGGCCGTGACATTGGTTTGTTGGTTGGTGGTTTTACCATGACAGCC 180

Qy     181 ACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCAGGTTAT 240
Db     181 ACCTGGGTTGGAGGAGGCTACATCAATGGGACAGCAGTAGCAGTGTATGGGCCAGGTTGT 240

Qy     241 GGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGTGGCCTG 300
Db     241 GGTCTAGCTTGGGCTCAGGCACCCATTGGATATTCTCTGAGTCTAATTTTAGGTGGTCTG 300

Qy     301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360
Db     301 TTTTTTGCGAAACCTATGCGTTCCAAGGGATATGTGACTATGTTAGACCCATTTCAACAG 360

Qy     361 ATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAAATGTTC 420
Db     361 ATCTATGGAAAGCGCATGGGTGGGCTGCTCTTCATCCCTGCACTGATGGGAGAGATGTTC 420

Qy     421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 480
Db     421 TGGGCTGCAGCAATTTTCTCTGCATTAGGGGCCACCATCAGCGTGATCATTGATGTGGAT 480

Qy     481 ATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTGGGAGGG 540
Db     481 GTGAACATATCGGTCATTGTCTCCGCACTCATTGCCATTCTTTATACCCTAGTGGGTGGG 540

Qy     541 CTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGGCTGTGG 600
Db     541 CTCTACTCTGTGGCATATACTGATGTTGTCCAGCTATTCTGCATTTTTATAGGACTGTGG 600

Qy     601 ATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACTGCTGTG 660
Db     601 ATCAGTGTCCCTTTTGCCCTGTCACATCCTGCAGTCACCGACATCGGATTCACAGCTGTG 660

Qy     661 CATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGG 720
Db     661 CATGCTAAATACCAGAGTCCCTGGCTGGGAACCATTGAATCAGTTGAAGTCTACACCTGG 720

Qy     721 CTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTCAGAGG 780
Db     721 CTTGATAATTTTCTGTTATTGATGCTGGGTGGAATCCCATGGCAAGCCTACTTCCAGAGG 780

Qy     781 GTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCTTTCGGG 840
Db     781 GTCCTCTCTTCATCCTCAGCCACCTATGCTCAGGTACTGTCCTACCTGGCAGCTTTTGGG 840

Qy     841 TGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACAGACTGG 900
Db     841 TGCCTGGTGATGGCTCTACCCGCCATATGCATAGGAGCTATTGGAGCTTCCACAGACTGG 900

Qy     901 AACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATGATTTTA 960
Db     901 AACCAGACTGCCTACGGGTATCCAGATCCCAAGACTAAGGAGGAAGCAGACATGATTCTC 960

Qy     961 CCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGTGCAGTT 1020
Db     961 CCGATCGTTCTGCAGTACCTCTGCCCTGTGTACATCTCCTTCTTTGGGCTTGGTGCTGTT 1020

Qy    1021 TCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATGTTTGCA 1080
Db    1021 TCAGCTGCTGTCATGTCCTCAGCTGACTCGTCCATCCTGTCGGCGAGTTCTATGTTTGCT 1080

Qy    1081 CGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140
Db    1081 CGGAATATCTACCAGCTTTCCTTCAGACAAAATGCATCAGACAAGGAAATTGTGTGGGTC 1140

Qy    1141 ATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTGACGAAA 1200
Db    1141 ATGAGGATCACTGTGCTTGTGTTCGGAGCATCTGCAACAGCCATGGCTTTGCTGACGAAG 1200

Qy    1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCCAG 1260
Db    1201 ACTGTGTATGGGCTGTGGTACCTGAGCTCTGACCTTGTCTACATCATCATCTTCCCACAG 1260

Qy    1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGTTATGTT 1320
Db    1261 CTGCTCTGTGTACTCTTCATCAAAGGAACCAACACTTATGGGGCAGTTGCTGGTTATATT 1320

Qy    1321 TCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCCTTGATC 1380
Db    1321 TTTGGACTATTCCTGAGAATTACTGGAGGAGAGCCATATCTATACTTGCAGCCCTTAATC 1380

Qy    1381 TTCTACCCTGGCTATTACCCTGATGATAATGGTATATATAATCAGAAATTTCCATTTAAA 1440
Db    1381 TTCTACCCTGGTTATTACTCTGACAAGAATGGTATATACAATCAGAGGTTCCCATTTAAA 1440

Qy    1441 ACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATCTCCTATCTAGCCAAGTAT 1500
Db    1441 ACTCTCTCCATGGTTACCTCATTCTTTACCAACATTTGTGTTTCTTATCTAGCCAAGTAT 1500

Qy    1501 CTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTTGATGCTGTTGTTGCAAGA 1560
Db    1501 CTATTTGAAAGTGGAACCTTGCCTCCAAAATTAGATGTATTTGATGCTGTTGTCGCAAGG 1560

Qy    1561 CACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAATGAAAATATTAAATTAGAT 1620
Db    1561 CACAGTGAAGAGAACATGGACAAGACCATTCTAGTCAGAAATGAAAATATCAAATTAAAT 1620
```

```
Qy   1621 GAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGCTCAACTTTCACCAATAAA 1680
          |||||||||| |||||| || || |||||| | |||||||| ||||||||||||||||||
Db   1621 GAACTTGCACCTGTGAAACCTCGGCAGAGCCTAACCCTCAGTTCAACTTTCACCAATAAG 1680

Qy   1681 GAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGGACTGAAGATAATTTACAG 1740
          |||||| |||||||||||||||||||||| || |||||||||||||||||||| |||||
Db   1681 GAGGCCCTCCTTGATGTTGATTCCAGTCCGGAGGGGTCTGGGACTGAAGATAACTTACAA 1740

Qy   1741 TGA 1743
          |||
Db   1741 TGA 1743
```

RESULT 14
TMA420808
LOCUS TMA420808 2528 bp mRNA linear VRT 27-NOV-2001
DEFINITION Torpedo marmorata mRNA for high affinity choline transporter (CHT1 gene).
ACCESSION AJ420808
VERSION AJ420808.1 GI:17148508
KEYWORDS CHT1 gene; high affinity choline transporter.
SOURCE marbled electric ray.
ORGANISM Torpedo marmorata
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes; Elasmobranchii; Squalea; Hypnosqualea; Pristiorajea; Batoidea; Torpediniformes; Torpedinoidei; Torpedinidae; Torpedo.
REFERENCE 1
AUTHORS Guermonprez,L., O'Regan,S., Meunier,F.M. and Morot-Gaudry-Talarmain,Y.
TITLE Cyclosporin, FK506 and rapamycin inhibit neuronal choline uptake via calcineurin-dependent and independent mechanisms
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2528)
AUTHORS O'Regan,S.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-2001) O'Regan S., Neurobiologie Cellulaire et Moleculaire, C.N.R.S., 1 av de la Terrasse, F-91198 Gif-sur-Yvette, FRANCE
FEATURES Location/Qualifiers
source 1. .2528
/organism="Torpedo marmorata"
/db_xref="taxon:7788"
/clone="tH312"
/tissue_type="electric lobe"
/tissue_lib="lambda ZAPII ELL"
gene 49. .2492
/gene="CHT1"
CDS 49. .1803
/gene="CHT1"
/function="neuronal Na-dependent choline transporter"
/codon_start=1
/evidence=experimental
/product="high affinity choline transporter"
/protein_id="CAD12727.1"
/db_xref="GI:17148509"
/translation="MTVHIDGIVAIVLFYLLILFVGLWAAWKSKNTSMEGAMDRSEAI
MIGGRDIGLLVGGFTMTATWVGGGYINGTAEAVYVPGYGLAWAQAPFGYALSLVIGGL
FFAKPMRSRGYVTMLDPFQQMYGKRMGGLLFIPALLGEIFWSAAILSALGATLSVIVD
ININVSVVVSAVIAVLYTLVGGLYSVAYTDVVQLFCIFLGLWISIPFALLNPAVTDII
VTANQEVYQEPWVGNIQSKDSLIWIDNFLLLMLGGIPWQVYFQRVLSASSATYAQVLS
FLAAFGCVLMAIPSVLIGAIGTSTDWNQTSYGLPGPIGKNETDMILPIVLQHLCPPYI
SFFGLGAVSAAVMSSADSSILSASSMFARNIYHLAFRQEASDKEIVWVMRITIFLFGG
AATSMALLAQSIYGLWYLSSDLVYVIIFPQLISVLFVKGTNTYGSIAGYIIGFLLRIS
GGEPYLHMQPFIYYPGCYLDHSFGDDPVYVQRFPFKTMAMLFSFLGNTGVSYLVKYLF
VSGILPPKLDFLDSVVSKHSKEIMDKTFLMNQDNITLSELVHVNPIHSASVSAALTNK
EAFEDIEPNPELSKSGND"
polyA_signal 2487. .2492
/gene="CHT1"
BASE COUNT 674 a 496 c 525 g 833 t
ORIGIN

Query Match 49.7%; Score 867; DB 5; Length 2528;
Best Local Similarity 69.3%; Pred. No. 9e-229;

Matches 1217; Conservative 0; Mismatches 520; Indels 18; Gaps 2;

```
Qy      1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCGTGTTCTACCTTCTAATTTTGCTG 60
          ||| |  | ||  | || ||  |  |||| ||| || |||| ||| |  |||| ||  |
Db     49 ATGACCGTTCACATCGATGGGATCGTAGCGATCGTCCTGTTTTACTTGTTAATCTTATTT 108

Qy     61 GTTGGAATATGGGCTGCCTGGAGAACCAAAAACAGTGGCA------GCGCAGAAGAGCGC 114
          |||||| |||||||||| |||| ||  |||||||     |       | |||   || ||
Db    109 GTTGGATTATGGGCTGCTTGGAAAAGTAAAAACACGTCAATGGAAGGAGCAATGGATCGG 168

Qy    115 AGCGAAGCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATG 174
          || ||||| || ||  |||| ||  ||||||| ||  |  |||||||||| || || |||
Db    169 AGTGAAGCTATAATGATTGGGGGAAGAGATATCGGGCTGCTGGTTGGTGGCTTCACAATG 228

Qy    175 ACAGCTACCTGGGTCGGAGGAGGGTATATCAATGGCACAGCTGAAGCAGTTTATGTACCA 234
          || || || |||||||| || || |||||||||| ||||| || || |||||||| ||
Db    229 ACCGCAACTTGGGTCGGTGGCGGTTATATCAATGGGACAGCAGAGGCGGTTTATGTTCCT 288

Qy    235 GGTTATGGCCTAGCTTGGGCTCAGGCACCAATTGGATATTCTCTTAGTCTGATTTTAGGT 294
          || || ||| | || ||||| ||||| ||  | |||||  | || || ||| || | ||
Db    289 GGGTACGGCTTGGCCTGGGCGCAGGCTCCCTTCGGATACGCACTCAGCCTGGTTATTGGC 348

Qy    295 GGCCTGTTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTT 354
          ||| | || || || ||||| ||||| ||| ||| || ||||||||| | |||||||||
Db    349 GGCTTATTTTTCGCTAAACCCATGCGCTCACGGGGTTACGTGACCATGCTGGACCCGTTT 408

Qy    355 CAGCAAATCTATGGAAAACGCATGGGCGGACTCCTGTTTATTCCTGCACTGATGGGAGAA 414
          || || || || || ||||| ||||| ||| | || || || || || ||  |||| |||
Db    409 CAACAGATGTACGGTAAACGAATGGGAGGATTGCTCTTCATCCCCGCTCTCCTGGGGGAA 468

Qy    415 ATGTTCTGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGAT 474
          || |||||| ||||||| ||  | || ||  | || || ||  | ||||||||  | ||
Db    469 ATCTTCTGGTCTGCAGCCATACTGTCCGCGCTAGGTGCAACTTTAAGCGTGATTGTGGAC 528

Qy    475 GTGGATATGCACATTTCTGTCATCATCTCTGCACTCATTGCCACTCTGTACACACTGGTG 534
          |  |||| || | || ||  |  | || ||  | || ||     | ||||| |||||
Db    529 ATCAATATAAACGTATCAGTGGTAGTTTCCGCTGTGATCGCTGTATTATACACTCTGGTC 588

Qy    535 GGAGGGCTCTATTCTGTGGCCTACACTGATGTCGTTCAGCTCTTTTGCATTTTTGTAGGG 594
          || ||| | || || || || ||||| |||||||| ||| | |||||||| ||  ||||
Db    589 GGCGGGTTATACTCGGTCGCGTACACAGATGTCGTCCAGTTGTTTTGCATCTTCTTAGGT 648

Qy    595 CTGTGGATCAGCGTCCCCTTTGCATTGTCACATCCTGCAGTCGCAGACATCGGGTTCACT 654
          |||||||||||| |||| || ||  | | |||| || ||  | || ||      | ||
Db    649 TTGTGGATCAGCATCCCTTTCGCCCTCTTAAATCCCGCTGTTACGGATATTATCGTGACC 708

Qy    655 GCTGTGCATGCCAAATACCAAAAGCCGTGGCTGGGAACTGTTGACTCATCTGAAGTCTAC 714
          ||    || |    || ||  |||| ||| |||||| | |  | ||| ||    |
Db    709 GCAAATCAAGAAGTTTATCAGGAGCCTTGGGTGGGAAATATACAATCAAAGGACAGTTTA 768

Qy    715 TCTTGGCTTGATAGTTTTCTGTTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTT 774
            ||| |||| |  ||||| ||  ||||||||||||||||||| ||||||| ||| |||
Db    769 ATCTGGATTGACAACTTTCTATTACTGATGCTGGGTGGAATCCCGTGGCAAGTATATTTT 828

Qy    775 CAGAGGGTTCTCTCTTCTTCCTCAGCCACCTATGCTCAAGTGCTGTCCTTCCTGGCAGCT 834
          ||||| || || ||| |||| || || |||||||| ||||| |||||||| ||||| ||
Db    829 CAGAGAGTCCTTTCTGCTTCTTCTGCTACCTATGCGCAAGTCCTGTCCTTTCTGGCTGCC 888

Qy    835 TTCGGGTGCCTGGTGATGGCCATCCCAGCCATACTCATTGGGGCCATTGGAGCATCAACA 894
          ||||| ||| |  | ||||||||||  |  | ||||| || || || ||| |||| ||
Db    889 TTCGGTTGCGTTCTCATGGCCATCCCGTCTGTTCTCATCGGTGCAATAGGAACATCCACT 948

Qy    895 GACTGGAACCAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGAGGCAGACATG 954
          |||||||| |||||| | ||||  | |||| || |  | | | ||| | || |||
Db    949 GACTGGAATCAGACTTCCTATGGCTTGCCAGGCCCTATAGGCAAAAATGAGACTGATATG 1008

Qy    955 ATTTTACCAATTGTTCTGCAGTATCTCTGCCCTGTGTATATTTCTTTCTTTGGTCTTGGT 1014
          ||||| || || || |||||| |||| || ||    || ||||| || ||||||||||
Db   1009 ATTTTGCCGATCGTGCTGCAGCATCTGTGTCCACCCTACATTTCCTTTTTTGGTCTTGGC 1068

Qy   1015 GCAGTTTCTGCTGCTGTTATGTCATCAGCAGATTCTTCCATCTTGTCAGCAAGTTCCATG 1074
          || || |||||||||| |||||||| || |||||||| ||||| |||||||||| |||
Db   1069 GCTGTCTCTGCTGCTGTGATGTCATCGGCTGATTCTTCTATCTTATCAGCAAGTTCTATG 1128
```

```
Qy   1075 TTTGCACGGAACATCTACCAGCTTTCCTTCAGACAAAATGCTTCGGACAAAGAAATCGTT 1134
          ||||| ||||| || ||||| ||| | ||||||||| | ||||| |||||||||||| ||
Db   1129 TTTGCTCGGAATATTTACCATCTTGCTTTCAGACAAGAGGCTTCAGACAAAGAAATAGTG 1188

Qy   1135 TGGGTTATGCGAATCACAGTGTTTGTGTTTGGAGCATCTGCAACAGCCATGGCCTTGCTG 1194
          ||||| ||||||||||  | ||| | ||||||| | |||||||| | ||||| |||||
Db   1189 TGGGTAATGCGAATCACCATATTTCTATTTGGAGGAGCTGCAACATCTATGGCATTGCTT 1248

Qy   1195 ACGAAAACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTTACATCGTTATCTTC 1254
           |  || |  | ||||| |||||||| || || || || ||||| ||| || |||||||
Db   1249 GCTCAATCAATCTATGGCCTCTGGTATCTGAGCTCAGATCTTGTCTACGTCATTATCTTT 1308

Qy   1255 CCCCAGCTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTGGCAGGT 1314
          || ||  |  | |  || ||||| || |||||||| ||||| ||||||  |  | || ||
Db   1309 CCTCAATTAATATCAGTGCTCTTCGTCAAGGGAACAAACACATATGGGTCTATTGCTGGA 1368

Qy   1315 TATGTTTCTGGCCTCTTCCTGAGAATAACTGGAGGGGAGCCATATCTGTATCTTCAGCCC 1374
          ||| |   |||| | || ||  | || | ||| || || |||||| |  || | |||||
Db   1369 TATATCATTGGCTTTTTGCTTCGGATTAGTGGTGGTGAACCATATTTACATATGCAGCCA 1428

     1375 TTGATCTTCTACCCTGGCT------------ATTACCCTGATGATAATGGTATATATAAT 1422
          || || |  |||||||| |             ||| |  ||  ||| ||  | | |||  |
Db   1429 TTTATTTATTACCCTGGATGCTATTTAGATCATTCCTTTGGAGATGATCCTGTTTATGTT 1488

Qy   1423 CAGAAATTTCCATTTAAAACACTTGCCATGGTTACATCATTCTTAACCAACATTTGCATC 1482
          |||| |||||| ||||||||  | || ||| |    || |||||   ||||| | |  |
Db   1489 CAGAGATTTCCCTTTAAAACCATGGCAATGTTATTCTCCTTCTTGGGCAACACTGGTGTA 1548

Qy   1483 TCCTATCTAGCCAAGTATCTATTTGAAAGTGGAACCTTGCCACCTAAATTAGATGTATTT 1542
          || ||||| | |||||| || || | ||||||||  |||||||| |||||||  |  ||
Db   1549 TCATATCTTGTCAAGTACCTGTTCGTAAGTGGAATATTGCCACCAAAATTAGACTTCCTT 1608

Qy   1543 GATGCTGTTGTTGCAAGACACAGTGAAGAAAACATGGATAAGACAATTCTTGTCAAAAAT 1602
          ||    |||||  ||| ||||||| | |||| |||||| || ||| |  |  | ||  |
Db   1609 GACAGCGTTGTATCAAAACACAGTAAGGAAATCATGGACAAAACATTCTTGATGAATCAG 1668

Qy   1603 GAAAATATTAAATTAGATGAACTTGCACTTGTGAAGCCACGACAGAGCATGACCCTCAGC 1662
          || |||||||  ||    || || |  | ||| || |||  ||| ||     |  | ||
Db   1669 GACAATATTACTTTGTCAGAGCTGGTGCATGTTAATCCAATACACAGTGCTTCAGTTAGT 1728

Qy   1663 TCAACTTTCACCAATAAAGAGGCCTTCCTTGATGTTGATTCCAGTCCAGAAGGGTCTGGG 1722
          |  |||| |||||||| || || ||    ||  ||||  | | ||| |||   |||  |
Db   1729 GCTGCTTTGACCAATAAGGAAGCATTTGAAGACATTGAGCCAAATCCTGAACTTTCTAAG 1788

Qy   1723 ACTGAAGATAATTTA 1737
          | |   || ||| |
Db   1789 TCAGGCAATGATTGA 1803
```

RESULT 15
HSA308384

```
LOCUS       HSA308384              2239 bp    DNA     linear   PRI 23-JAN-2002
DEFINITION  Homo sapiens partial CHT1 gene for high affinity choline
            transporter, exon 9.
ACCESSION   AJ308384
VERSION     AJ308384.1  GI:18369775
KEYWORDS    ChT1 gene; high affinity choline transporter.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Wieland,A., Bonisch,H. and Bruss,M.
  TITLE     Molecular cloning of the human and murine high affinity choline
            transporters and characterizationof the human gene structure
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2239)
  AUTHORS   Bruess,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JAN-2001) Bruess M., University of Bonn, Pharmacology
            and Toxicology, Reuterstr. 2b, D-53113 Bonn, GERMANY
FEATURES             Location/Qualifiers
     source          1. .2239
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q11-13"
     gene            1. .2239
                     /gene="CHT1"
     intron          <1. .791
                     /gene="CHT1"
                     /number=8
     exon            792. .>2239
                     /gene="CHT1"
                     /number=9
                     /usedin=AJ308378:CHT1_CDS
                     /usedin=AJ308377:CHT1_mRNA
BASE COUNT      664 a    442 c    440 g    693 t
ORIGIN
```

Query Match 36.2%; Score 630.8; DB 9; Length 2239;
Best Local Similarity 99.7%; Pred. No. 2.4e-163;
Matches 632; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy   1110 AAATGCTTCGGACAAAGAAATCGTTTGGGTTATGCGAATCACAGTGTTTGTGTTTGGAGC 1169
          | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    788 ACAGGCTTCGGACAAAGAAATCGTTTGGGTTATGCGAATCACAGTGTTTGTGTTTGGAGC 847

Qy   1170 ATCTGCAACAGCCATGGCCTTGCTGACGAAAACTGTGTATGGGCTCTGGTACCTCAGTTC 1229
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    848 ATCTGCAACAGCCATGGCCTTGCTGACGAAAACTGTGTATGGGCTCTGGTACCTCAGTTC 907

Qy   1230 TGACCTTGTTTACATCGTTATCTTCCCCCAGCTGCTTTGTGTACTCTTTGTTAAGGGAAC 1289
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    908 TGACCTTGTTTACATCGTTATCTTCCCCCAGCTGCTTTGTGTACTCTTTGTTAAGGGAAC 967

Qy   1290 CAACACCTATGGGGCCGTGGCAGGTTATGTTTCTGGCCTCTTCCTGAGAATAACTGGAGG 1349
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    968 CAACACCTATGGGGCCGTGGCAGGTTATGTTTCTGGCCTCTTCCTGAGAATAACTGGAGG 1027

Qy   1350 GGAGCCATATCTGTATCTTCAGCCCTTGATCTTCTACCCTGGCTATTACCCTGATGATAA 1409
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1028 GGAGCCATATCTGTATCTTCAGCCCTTGATCTTCTACCCTGGCTATTACCCTGATGATAA 1087

Qy   1410 TGGTATATATAATCAGAAATTTCCATTTAAAACACTTGCCATGGTTACATCATTCTTAAC 1469
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1088 TGGTATATATAATCAGAAATTTCCATTTAAAACACTTGCCATGGTTACATCATTCTTAAC 1147

Qy   1470 CAACATTTGCATCTCCTATCTAGCCAAGTATCTATTTGAAAGTGGAACCTTGCCACCTAA 1529
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1148 CAACATTTGCATCTCCTATCTAGCCAAGTATCTATTTGAAAGTGGAACCTTGCCACCTAA 1207

Qy   1530 ATTAGATGTATTTGATGCTGTTGTTGCAAGACACAGTGAAGAAAACATGGATAAGACAAT 1589
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1208 ATTAGATGTATTTGATGCTGTTGTTGCAAGACACAGTGAAGAAAACATGGATAAGACAAT 1267

Qy   1590 TCTTGTCAAAAATGAAAATATTAAATTAGATGAACTTGCACTTGTGAAGCCACGACAGAG 1649
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1268 TCTTGTCAAAAATGAAAATATTAAATTAGATGAACTTGCACTTGTGAAGCCACGACAGAG 1327

Qy   1650 CATGACCCTCAGCTCAACTTTCACCAATAAAGAGGCCTTCCTTGATGTTGATTCCAGTCC 1709
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1328 CATGACCCTCAGCTCAACTTTCACCAATAAAGAGGCCTTCCTTGATGTTGATTCCAGTCC 1387

Qy   1710 AGAAGGGTCTGGGACTGAAGATAATTTACAGTGA 1743
          |||||||||||||||||||||||||||||||||
Db   1388 AGAAGGGTCTGGGACTGAAGATAATTTACAGTGA 1421
```

Search completed: March 31, 2003, 17:49:36
Job time : 4580 secs

OM protein - protein search, using sw model

Run on: March 31, 2003, 16:34:01 ; Search time 67 Seconds
(without alignments)
1153.513 Million cell updates/sec

Title: US-10-069-541-6
Perfect score: 2972
Sequence: 1 MAFHVEGLIAIIVFYLLILL.....EAFLDVDSSPEGSGTEDNLQ 580

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2972	100.0	580	22	AAB74665	Human high affinit
2	2972	100.0	580	22	AAB86837	Human CHOT protein
3	2820	94.9	580	22	AAB74664	Rat high affinity
4	2810	94.5	580	22	AAY72388	Mouse P4P6B1 OMA (
5	2801	94.2	580	22	AAB74666	Mouse high affinit
6	1453	48.9	576	22	AAB74663	C. elegans high af
7	723.5	24.3	278	22	ABB71601	Drosophila melanog
8	344	11.6	492	22	AAB96805	Putative P. abyssi
9	311.5	10.5	675	22	AAB85102	Novel human transp
10	310.5	10.4	675	22	AAB60093	Human transport pr
11	310.5	10.4	675	23	ABB80599	Human sbg1007026SG
12	308.5	10.4	662	16	AAR73595	Cotransporter prot
13	306	10.3	660	22	ABG30151	Novel human diagno
14	298.5	10.0	675	23	AAU77134	Human sodium-sugar
15	298	10.0	672	16	AAR73593	Cotransporter prot
16	287	9.7	659	20	AAY31221	Human SAAT1 protei
17	283	9.5	484	22	AAB96255	Putative P. abyssi
18	277.5	9.3	596	22	AAE08088	Human transporter-
19	277.5	9.3	596	22	AAE06614	Human protein havi
20	277.5	9.3	596	23	ABB80588	Human sbg1020829SG
21	277.5	9.3	596	23	AAO14199	Human transporter
22	272.5	9.2	524	22	AAG93195	C glutamicum prote
23	272.5	9.2	524	22	AAB76757	Corynebacterium gl
24	272.5	9.2	718	22	ABG01897	Novel human diagno
25	272.5	9.2	718	23	AAB47976	BCW2. Homo sapien
26	272.5	9.2	718	23	AAG78575	Human SLC5A3 amino
27	271	9.1	537	22	AAB96626	Putative P. abyssi
28	262.5	8.8	518	23	ABP39899	Staphylococcus epi
29	260	8.7	681	23	AAO14202	Human transporter
30	260	8.7	752	22	ABG28100	Novel human diagno
31	260	8.7	752	23	AAE16783	Human transporter
32	248	8.3	742	23	AAE16778	Human transporter
33	232.5	7.8	465	20	AAY35106	Chlamydia pneumoni
34	231	7.8	643	22	AAB73920	Human NIS. Homo s
35	230.5	7.8	664	23	AAU85407	Human protein NOV5
36	222	7.5	587	22	ABB62630	Drosophila melanog
37	222	7.5	591	23	ABP61007	Novel human protei
38	219	7.4	551	22	AAG90677	C glutamicum prote
39	219	7.4	551	22	AAB76764	Corynebacterium gl
40	218	7.3	618	18	AAW32105	Rat sodium/iodide
41	218	7.3	618	22	AAB73921	Rat NIS. Rattus s
42	216.5	7.3	618	22	AAE06584	Human protein havi
43	216.5	7.3	618	23	AAO14195	Human transporter
44	209	7.0	635	21	AAY71060	Human membrane tra
45	206.5	6.9	530	23	ABP39665	Staphylococcus epi

ALIGNMENTS

RESULT 1
AAB74665
ID AAB74665 standard; Protein; 580 AA.
XX
AC AAB74665;
XX
DT 01-JUN-2001 (first entry)
XX
DE Human high affinity choline transporter protein.
XX
KW High affinity choline transporter; cho-1; Alzheimer's disease;
KW diagnosis.
XX
OS Homo sapiens.
XX
PN WO200116315-A1.
XX
PD 08-MAR-2001.
XX
PF 18-AUG-2000; 2000WO-JP05545.
XX
PR 27-AUG-1999; 99JP-0240642.
PR 27-DEC-1999; 99JP-0368991.
XX
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI Haga T, Okuda T;
XX
DR WPI; 2001-226688/23.
DR N-PSDB; AAF81712.
XX
PT New rat and human spinal cord high affinity choline transporters,
PT useful in diagnosis of Alzheimer's disease and screening promoters as

PT drugs for treating Alzheimer's disease -
XX
PS Claim 8; Page 76-78; 90pp; Japanese.
XX
CC The present sequence represents a human (Homo sapiens) high affinity
CC choline transporter protein designated cho-1. The cho-1 protein has
CC nootropic and neuroprotective activities. The cho-1 polynucleotide and
CC protein can be used for the diagnosis of diseases related to the
CC expression of cho-1 by comparing the cho-1 polynucleotide sequence in a
CC sample to that of a control. Drug compositions containing the cho-1
CC protein or expression promoters or inhibitors of cho-1 are useful for
CC treating disorders characterised by abnormal levels of cho-1, such as
CC Alzheimer's disease.
XX
SQ Sequence 580 AA;

Query Match 100.0%; Score 2972; DB 22; Length 580;
Best Local Similarity 100.0%; Pred. No. 2.3e-289;
Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60

Qy     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120

Qy    121 IYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 IYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180

Qy    181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSW 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSW 240

Qy    241 LDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDW 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 LDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDW 300

Qy    301 NQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 NQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360

Qy    361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420

Qy    421 LLCVLFVKGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 LLCVLFVKGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480

Qy    481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAVVARHSEENMDKTILVKNENIKLD 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAVVARHSEENMDKTILVKNENIKLD 540

Qy    541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
          |||||||||||||||||||||||||||||||||||||||||
Db    541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
```

RESULT 2
AAB86837
ID AAB86837 standard; Protein; 580 AA.
XX
AC AAB86837;
XX
DT 26-NOV-2001 (first entry)
XX
DE Human CHOT protein.
XX
KW CHOT; human; choline transporter; chromosome 2q11-13; nootropic;
KW neuroprotective; gene therapy; antisense therapy; degenerative disease;
KW cognitive disorder; Alzheimer's disease.
XX
OS Homo sapiens.
XX
PN DE10009055-A1.
XX
PD 30-AUG-2001.
XX
PF 28-FEB-2000; 2000DE-1009055.
XX
PR 28-FEB-2000; 2000DE-1009055.
XX
PA (BRUE/) BRUESS M.
PA (BOEN/) BOENISCH H.
XX
PI Bruess M, Boenisch H;
XX
DR WPI; 2001-590709/67.
DR N-PSDB; AAH49207.
XX
PT A new gene encoding human choline transporter, designated hCHOT is
PT located on chromosome 2q11-13 and is useful to treat degenerative
PT disorders such as Alzheimer's disease -
XX
PS Disclosure; Page 11; 12pp; German.
XX
CC This invention describes a novel gene encoding human choline transporter,
CC designated hCHOT which is located on chromosome 2q11-13. The products of
CC the invention have nootropic and neuroprotective activity and can be used
CC for gene or antisense therapy. (I) is used to treat degenerative disease,
CC particularly cognitive disorders such as Alzheimer's disease. Sense and
CC antisense oligonucleotides derived from the gene may be used in
CC diagnostics and other techniques. This sequence represents the human CHOT
CC protein described in the invention.
XX
SQ Sequence 580 AA;

Query Match 100.0%; Score 2972; DB 22; Length 580;
Best Local Similarity 100.0%; Pred. No. 2.3e-289;
Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60

Qy     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120

Qy    121 IYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 IYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180

Qy    181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSW 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSW 240

Qy    241 LDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDW 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 LDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDW 300

Qy    301 NQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 NQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360

Qy    361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420

Qy    421 LLCVLFVKGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 LLCVLFVKGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480
```

```
Qy    481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAVVARHSEENMDKTILVKNENIKLD 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAVVARHSEENMDKTILVKNENIKLD 540

Qy    541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
          |||||||||||||||||||||||||||||||||||||||||
Db    541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580


RESULT 3
AAB74664
ID   AAB74664 standard; Protein; 580 AA.
XX
AC   AAB74664;
XX
DT   01-JUN-2001  (first entry)
XX
DE   Rat high affinity choline transporter protein.
XX
KW   High affinity choline transporter; cho-1; Alzheimer's disease;
KW   diagnosis.
XX
OS   Rattus norvegicus.
XX
PN   WO200116315-A1.
XX
PD   08-MAR-2001.
XX
PF   18-AUG-2000; 2000WO-JP05545.
XX
PR   27-AUG-1999;   99JP-0240642.
PR   27-DEC-1999;   99JP-0368991.
XX
PA   (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI   Haga T,  Okuda T;
XX
DR   WPI; 2001-226688/23.
DR   N-PSDB; AAF81711.
XX
PT   New rat and human spinal cord high affinity choline transporters,
PT   useful in diagnosis of Alzheimer's disease and screening promoters as
PT   drugs for treating Alzheimer's disease -
XX
PS   Claim 5; Page 69-71; 90pp; Japanese.
XX
CC   The present sequence represents a rat (Rattus norvegicus) high affinity
CC   choline transporter protein designated cho-1. The cho-1 protein has
CC   nootropic and neuroprotective activities. The cho-1 polynucleotide and
CC   protein can be used for the diagnosis of diseases related to the
CC   expression of cho-1 by comparing the cho-1 polynucleotide sequence in a
CC   sample to that of a control. Drug compositions containing the cho-1
CC   protein or expression promoters or inhibitors of cho-1 are useful for
CC   treating disorders characterised by abnormal levels of cho-1, such as
CC   Alzheimer's disease.
XX
SQ   Sequence   580 AA;

  Query Match             94.9%;  Score 2820;  DB 22;  Length 580;
  Best Local Similarity   93.1%;  Pred. No. 4.3e-274;
  Matches  540;  Conservative   24;  Mismatches   16;  Indels    0;  Gaps    0;

Qy      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
          | ||||||:|||:||||| ||||||||:||||:|||||||||||||||||||||||||||
Db      1 MPFHVEGLVAIILFYLLIFLVGIWAAWKTKNSGNAEERSEAIIVGGRDIGLLVGGFTMTA 60

Qy     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
          ||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db     61 TWVGGGYINGTAEAVYGPGCGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120

Qy    121 IYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180
          ||||||||||||||||||||||||||||||||||||||::||||:|||||| |||||||
Db    121 IYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDVNISVIVSALIAILYTLVGG 180

Qy    181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSW 240
          |||||||||||||||:|||||||||||||| |||||||||| |||||::| |||:|
Db    181 LYSVAYTDVVQLFCIFIGLWISVPFALSHPAVTDIGFTAVHAKYQSPWLGTIESVEVYTW 240

Qy    241 LDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDW 300
          ||:|||||||||||||||||||||||||||||||||||||||||:||| ||||||||||
Db    241 LDNFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMALPAICIGAIGASTDW 300

Qy    301 NQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360
          |||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 NQTAYGFPDPKTKEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360

Qy    361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||
Db    361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIIIFPQ 420

Qy    421 LLCVLFVKGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480
          |||||:|||||||||||: ||||||||||||||||||||||||||| ||||||:||||
Db    421 LLCVLFIKGTNTYGAVAGYIFGLFLRITGGEPYLYLQPLIFYPGYYPDKNGIYNQRFPFK 480

Qy    481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAVVARHSEENMDKTILVKNENIKLD 540
          ||:||||| |||||||||||||||||||:|||||:|||||||||||||:||||||:
Db    481 TLSMVTSFFTNICVSYLAKYLFESGTLPPKLDIFDAVVSRHSEENMDKTILVRNENIKLN 540

Qy    541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
          ||| ||||||:|||||||||||| ||||||||||||||||
Db    541 ELAPVKPRQSLTLSSTFTNKEALLDVDSSPEGSGTEDNLQ 580


RESULT 4
AAY72388
ID   AAY72388 standard; Protein; 580 AA.
XX
AC   AAY72388;
XX
DT   24-APR-2001  (first entry)
XX
DE   Mouse P4P6B1 OMA (obese mice adipocyte) protein.
XX
KW   Mouse; OMA protein; obese mice adipocyte; P4P6B1;
KW   fuel metabolism disorder; therapy; obesity; diabetes; gene therapy;
KW   anorectic; antidiabetic.
XX
OS   Mus sp.
XX
PN   WO200078950-A2.
XX
PD   28-DEC-2000.
XX
PF   13-JUN-2000; 2000WO-US16217.
XX
PR   22-JUN-1999;   99US-0141515.
XX
PA   (AMYL-) AMYLIN PHARM INC.
XX
PI   Sierzega M,  Albrandt K;
XX
DR   WPI; 2001-112322/12.
DR   N-PSDB; AAD02457.
XX
PT   Novel obese mice adipocyte polypeptides useful in diagnosis and
PT   treatment of disorders of fuel metabolism such as obesity or diabetes
PT   -
XX
PS   Claim 11; Fig 3; 83pp; English.
XX
CC   The present sequence is mouse OMA (obese mice adipocyte) protein
CC   encoded by P4P6B1 cDNA. The P4P6B1 cDNA fragment was generated by
CC   RNA fingerprinting using random primers P4 and P6.
CC   OMA is used as a diagnostic reagent for diagnosing a disorder of fuel
CC   metabolism in an underweight or an overweight individual, by detecting
CC   the transcription level of a gene encoding OMA, which is induced or
```

```
CC   repressed in an individual by a factor such as genetic obesity, fasting
CC   and refeeding of a fasted individual. OMA is useful in the generation of
CC   antibodies, for use in pharmaceutical compositions and for studying
CC   DNA/protein interactions. Nucleic acids encoding OMA are involved in
CC   gene therapy. An inhibitor of OMA or an antisense oligonucleotide that
CC   inhibits expression of OMA are useful for treating disorders of fuel
CC   metabolism such as obesity or diabetes.
XX
SQ   Sequence   580 AA;

  Query Match             94.5%;  Score 2810;  DB 22;  Length 580;
  Best Local Similarity   93.1%;  Pred. No. 4.3e-273;
  Matches  540;  Conservative   23;  Mismatches   17;  Indels    0;  Gaps    0;

Qy       1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
           |:|||||:|||:|||||||||||||||:|||||: ||||||||||||||||||||||||||
Db       1 MSFHVEGLVAIILFYLLILLVGIWAAWKTKNSGNPEERSEAIIVGGRDIGLLVGGFTMTA 60

Qy      61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
           |||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||
Db      61 TWVGGGYINGTAEAVYGPGCGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120

Qy     121 IYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180
           ||||||||||||||||||||||||||||||||||||||::||||:||||| |||||||
Db     121 IYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDVNISVIVSALIAILYTLVGG 180

Qy     181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSW 240
           ||||||||||||||:|||||||||||| |||||||||| |||||::| |||:|
Db     181 LYSVAYTDVVQLFCIFIGLWISVPFALSHPAVTDIGFTAVHAKYQSPWLGTIESVEVYTW 240

Qy     241 LDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDW 300
           ||:||||||||||||||||||||||||||||||||||||||||:||| ||||||||||
Db     241 LDNFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMALPAICIGAIGASTDW 300

Qy     301 NQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360
           |||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 NQTAYGYPDPKTKEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360

Qy     361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||:||||
Db     361 RNIYQLSFRQNASDKEIVWVMRITVLVFGASATAMALLTKTVYGLWYLSSDLVYIIIFPQ 420

Qy     421 LLCVLFVKGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480
           |||||:|||||||||||: |||||||||||||||||||||||| | ||||||:||||
Db     421 LLCVLFIKGTNTYGAVAGYIFGLFLRITGGEPYLYLQPLIFYPGYYSDKNGIYNQRFPFK 480

Qy     481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAVVARHSEENMDKTILVKNENIKLD 540
           ||:||||| |||:|||||||||||||||||||||||||||||||||||||||||||||:
Db     481 TLSMVTSFFTNICVSYLAKYLFESGTLPPKLDVFDAVVARHSEENMDKTILVRNENIKLN 540

Qy     541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
           ||| ||||||:|||||||||| ||||||||||||||||
Db     541 ELAPVKPRQSLTLSSTFTNKEALLDVDSSPEGSGTEDNLQ 580


RESULT 5
AAB74666
ID   AAB74666 standard; Protein; 580 AA.
XX
AC   AAB74666;
XX
DT   01-JUN-2001  (first entry)
XX
DE   Mouse high affinity choline transporter protein.
XX
KW   High affinity choline transporter; cho-1; Alzheimer's disease;
KW   diagnosis.
XX
OS   Mus musculus.
XX
PN   WO200116315-A1.
XX
PD   08-MAR-2001.
XX
PF   18-AUG-2000; 2000WO-JP05545.
XX
PR   27-AUG-1999;   99JP-0240642.
PR   27-DEC-1999;   99JP-0368991.
XX
PA   (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI   Haga T,  Okuda T;
XX
DR   WPI; 2001-226688/23.
DR   N-PSDB; AAF81713.
XX
PT   New rat and human spinal cord high affinity choline transporters,
PT   useful in diagnosis of Alzheimer's disease and screening promoters as
PT   drugs for treating Alzheimer's disease -
XX
PS   Claim 11; Page 82-85; 90pp; Japanese.
XX
CC   The present sequence represents a mouse (Mus musculus) high affinity
CC   choline transporter protein designated cho-1. The cho-1 protein has
CC   nootropic and neuroprotective activities. The cho-1 polynucleotide and
CC   protein can be used for the diagnosis of diseases related to the
CC   expression of cho-1 by comparing the cho-1 polynucleotide sequence in a
CC   sample to that of a control. Drug compositions containing the cho-1
CC   protein or expression promoters or inhibitors of cho-1 are useful for
CC   treating disorders characterised by abnormal levels of cho-1, such as
CC   Alzheimer's disease.
XX
SQ   Sequence   580 AA;

  Query Match             94.2%;  Score 2801;  DB 22;  Length 580;
  Best Local Similarity   92.8%;  Pred. No. 3.5e-272;
  Matches  538;  Conservative   23;  Mismatches   19;  Indels    0;  Gaps    0;

Qy       1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
           |:|||||:|||:||||| ||||||||:|||||: || ||||||||||||||||||||||
Db       1 MSFHVEGLVAIILFYLLIFLVGIWAAWKTKNSGNPEEHSEAIIVGGRDIGLLVGGFTMTA 60

Qy      61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
           |||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||
Db      61 TWVGGGYINGTAEAVYGPGCGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120

Qy     121 IYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180
           ||||||||||||||||||||||||||||||||||||||::||||:||||| |||||||
Db     121 IYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDVNISVIVSALIAILYTLVGG 180

Qy     181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSW 240
           ||||||||||||||:|||||||||||| |||||||||| |||||::| |||:|
Db     181 LYSVAYTDVVQLFCIFIGLWISVPFALSHPAVTDIGFTAVHAKYQSPWLGTIESVEVYTW 240

Qy     241 LDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDW 300
           ||:||||||||||||||||||||||||||||||||||||||||:||| ||||||||||
Db     241 LDNFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMALPAICIGAIGASTDW 300

Qy     301 NQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360
           |||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 NQTAYGYPDPKTKEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360

Qy     361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||:||||
Db     361 RNIYQLSFRQNASDKEIVWVMRITVLVFGASATAMALLTKTVYGLWYLSSDLVYIIIFPQ 420

Qy     421 LLCVLFVKGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480
           |||||:|||||||||||: |||||||||||||||||||||||| | ||||||:||||
Db     421 LLCVLFIKGTNTYGAVAGYIFGLFLRITGGEPYLYLQPLIFYPGYYSDKNGIYNQRFPFK 480

Qy     481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAVVARHSEENMDKTILVKNENIKLD 540
           ||:||||| |||:|||||||||||||||||||||||||||||||||||||||||||||:
Db     481 TLSMVTSFFTNICVSYLAKYLFESGTLPPKLDVFDAVVARHSEENMDKTILVRNENIKLN 540

Qy     541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
           ||| ||||||:|||||||||| ||||||||||||||||
```

```
Db    541 ELAPVKPRQSLTLSSTFTNKEALLDVDSSPEGSGTEDNLQ 580


RESULT 6
AAB74663
ID   AAB74663 standard; Protein; 576 AA.
XX
AC   AAB74663;
XX
DT   01-JUN-2001  (first entry)
XX
DE   C. elegans high affinity choline transporter protein.
XX
KW   High affinity choline transporter; cho-1; Alzheimer's disease;
KW   diagnosis.
XX
OS   Caenorhabditis elegans.
XX
PN   WO200116315-A1.
XX
PD   08-MAR-2001.
XX
PF   18-AUG-2000; 2000WO-JP05545.
XX
PR   27-AUG-1999;   99JP-0240642.
PR   27-DEC-1999;   99JP-0368991.
XX
PA   (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI   Haga T,  Okuda T;
XX
DR   WPI; 2001-226688/23.
DR   N-PSDB; AAF81710.
XX
PT   New rat and human spinal cord high affinity choline transporters,
PT   useful in diagnosis of Alzheimer's disease and screening promoters as
PT   drugs for treating Alzheimer's disease -
XX
PS   Claim 2; Page 62-64; 90pp; Japanese.
XX
CC   The present sequence represents a Caenorhabditis elegans high affinity
CC   choline transporter protein designated cho-1. The cho-1 protein has
CC   nootropic and neuroprotective activities. The cho-1 polynucleotide and
CC   protein can be used for the diagnosis of diseases related to the
CC   expression of cho-1 by comparing the cho-1 polynucleotide sequence in a
CC   sample to that of a control. Drug compositions containing the cho-1
CC   protein or expression promoters or inhibitors of cho-1 are useful for
CC   treating disorders characterised by abnormal levels of cho-1, such as
CC   Alzheimer's disease.
XX
SQ   Sequence   576 AA;

  Query Match           48.9%;  Score 1453;  DB 22;  Length 576;
  Best Local Similarity  50.5%;  Pred. No. 8.6e-137;
  Matches  295;  Conservative   95;  Mismatches  150;  Indels   44;  Gaps    9;

Qy       7 GLIAIIVFYLLILLVGIWAAWRTKNSGSAEER----SEAIIVGGRDIGLLVGGFTMTATW 62
           |::||: ||:|||:|||||  ::|:|   |       :| ::: ||:|| ||| |||||||
Db       6 GIVAIVFFYVLILVVGIWAGRKSKSSKELESEAGAATEEVMLAGRNIGTLVGIFTMTATW 65

Qy      63 VGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQIY 122
           ||| ||||||||:|    ||   |||:||::||::||| ||| || :||:|||||||  |
Db      66 VGGAYINGTAEALY--NGGLLGCQAPVGYAISLVMGGLLFAKKMREEGYITMLDPFQHKY 123

Qy     123 GKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGGLY 182
           |:|:|||:::|||:|| || ||| ||||||:|||: :||: || :|| ||  ||  || |
Db     124 GQRIGGLMYVPALLGETFWTAAILSALGATLSVILGIDMNASVTLSACIAVFYTFTGGYY 183

Qy     183 SVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDS-SEVYSWL 241
           :|||||||||||||||: || |: |    ||   |        |:| :    |   |:
Db     184 AVAYTDVVQLFCIFVGLWVCVPAAMVHDGAKDISRNA------GDWIGEIGGFKETSLWI 237

Qy     242 DSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWN 301
           |  |||: |||||| |||||||| :|  || |||:|  ||::|||| |||||  :|||
Db     238 DCMLLLVFGGIPWQVYFQRVLSSKTAHGAQTLSFVAGVGCILMAIPPALIGAIARNTDWR 297

Qy     302 QTAYGLPDPKTTEEA------DMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSA 355
             | |   :  |  |:        :|::|:| ||| | :::| ||||||||||||||:|||
Db     298 MTDYSPWNNGTKVESIPPDKRNMVVPLVFQYLTPRWVAFIGLGAVSAAVMSSADSSVLSA 357

Qy     356 SSMFARNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYI 415
           :|||| ||::|: | :||:||:: |||| :   |  || |||  :::|||||| :|||||:
Db     358 ASMFAHNIWKLTIRPHASEKEVIIVMRIAIICVGIMATIMALTIQSIYGLWYLCADLVYV 417

Qy     416 VIFPQLLCVLFVKGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQ 475
           ::|||||||:::  :||||::|||  || ||: |||| : |     || |    :|:  |
Db     418 ILFPQLLCVVYMPRSNTYGSLAGYAVGLVLRLIGGEPLVSLPAFFHYPMY---TDGV--Q 472

Qy     476 KFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAVV---ARHSEENMDKTILV 532
            |||:| ||::|  |   :|  :: ||:|| | |: ||   ||     |    | :  |
Db     473 YFPFRTTAMLSSMATIYIVSIQSEKLFKSGRLSPEWDVMGCVVNIPIDHVPLPSDVSFAV 532

Qy     533 KNENIKL----------------DELALVKPRQSMTLSSTFTN 559
           :| : :                  ||  |: |    :  || :|
Db     533 SSETLNMKAPNGTPAPVHPNQQPSDENTLLHPYSDQSYYSTNSN 576


RESULT 7
ABB71601
ID   ABB71601 standard; Protein; 278 AA.
XX
AC   ABB71601;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 41595.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL15704.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 41595; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
```

SQ Sequence 278 AA;

Query Match 24.3%; Score 723.5; DB 22; Length 278;
Best Local Similarity 64.2%; Pred. No. 5.5e-64;
Matches 136; Conservative 33; Mismatches 38; Indels 5; Gaps 3;

```
Qy       4 HVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWV 63
           :: |:::|::|||||||:|||||  | | ||:  |  | ::: || ||| || |||||||
Db       3 NIAGVVSIVLFYLLILVVGIWAG-RKKQSGNDSE--EEVMLAGRSIGLFVGIFTMTATWV 59

Qy      64 GGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQIYG 123
           ||||||||||||:|    || | ||| ||:|||: ||:||| ||| :||:|||||  |  :|
Db      60 GGGYINGTAEAIYTS--GLVWCQAPFGYALSLVFGGIFFANPMRKQGYITMLDPLQDSFG 117

Qy     124 KRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGGLYS 183
           :|||||||:||| ||:|||| | :|||||:|||||:|   |||:|: ||  ||| |||||
Db     118 ERMGGLLFLPALCGEVFWAAGILAALGATLSVIIDMDHRTSVILSSCIAIFYTLFGGLYS 177

Qy     184 VAYTDVVQLFCIFVGLWISVPFALSHPAVADI 215
           ||||||:||||||:|||: :||| |:  |  :
Db     178 VAYTDVIQLFCIFIGLWMCIPFAWSNEHVGSL 209
```

RESULT 8
AAB96805
ID AAB96805 standard; Protein; 492 AA.
XX
AC AAB96805;
XX
DT 29-OCT-2001 (first entry)
XX
DE Putative P. abyssi permease #30.
XX
KW Hyperthermophilic archaeon; hyperthermophilic protein.
XX
OS Pyrococcus abyssi.
XX
PN FR2792651-A1.
XX
PD 27-OCT-2000.
XX
PF 21-APR-1999; 99FR-0005034.
XX
PR 21-APR-1999; 99FR-0005034.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
PA (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
PI Querellou J, Weissenbach J, Saurin W, Heilig R;
XX
DR WPI; 2001-126236/14.
XX
PT New nucleotide sequences isolated from Pyrococcus abyssi encode
PT proteins useful in industry -
XX
PS Claim 7; Pages 1579-1580; 1657pp; French.
XX
CC The present invention relates to the genomic sequence of Pyrococcus
CC abyssi (see AAF86431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
CC a hyperthermophilic archaeon, which is isolated from deep-sea
CC hydrothermal vents. The present sequence is one such P. abyssi protein.
CC The proteins of the present invention have various potential industrial
CC uses, since the proteins are stable at very high temperatures, some up to
CC 110 degrees centigrade.
CC Note: This patent is in the same patent family as WO200065062, which
CC contains additional sequences as shown in AAB99132-AAB99143,
CC AAH75903-AAH75920 and AAG66436.
XX
SQ Sequence 492 AA;

Query Match 11.6%; Score 344; DB 22; Length 492;
Best Local Similarity 24.2%; Pred. No. 1.7e-25;
Matches 132; Conservative 99; Mismatches 196; Indels 118; Gaps 25;

```
Qy       8 LIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY 67
           |:| :  :| :||| :|  |||:        |   |||| :  :    :  ::
Db      14 LVAFLFTLILPILVGFYAMKRTKS-------EEDFFVGGRAMDKITVALSAVSSGRSSWL 66

Qy      68 INGTAEAVYVPGYGLAWAQAPIGYSLS----LILGGLFFAKPMRSKGYVTMLDPFQQIYG 123
           : | :   |  |    |  | :|| ::     :  |:   |       :|: | |:  :
Db      67 VLGLSGMAYKMGVTAVW--AAVGYIVAEMFQFVYMGIRLRKFSERFNAITVPDYFEARFR 124

Qy     124 K-----RMGG----LLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATL 174
                  |:       ::|: : :|  | | |       |:|  : : :  :::|| |:  :
Db     125 DTSKILRIAASIIIIIFLTSYVGAQFNAGA------KTLSTALGISIFTALMISVLMIIV 178

Qy     175 YTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFT--------AVHAKYQK 226
           | ::||  :||| ||::   : :|| :     |   ||| :| |         |:  |
Db     179 YMILGGFIAVAYNDVIRAVIMIIGLVV-----LPVIAVAKVGGTEEVLKVLHALDPKLIN 233

Qy     227 PW---LGTVDSSEVYSWLDSFLLLMLG-GIPWQAY-FQRVLSSSSATYAQVLSFLAAFGC 281
           ||    | |                | :| | | |  :  | :|       :| : :  |
Db     234 PWAFGAGVVIG----------FLGIGFGSPGQPHIIVRYMSIDDPNKLRVSTVVGTFWN 282

Qy     282 LVMAIPAILIGAIGASTDWNQTAYGLPDPKTT--EEADMILP-IVLQYLCPVYISFFGLG 338
           :|:|  || :|  | :          :||      : |:|| | :  ||  |:      :|
Db     283 VVLAWGAIFVGLAGRAI--------VPDVSQLPGKNAEMIYPYLSAQYFPPILYGIL-IG 333

Qy     339 AVSAAVMSSADSSILSASSMFARNIYQLSFRQNA--SDKEIVWVMRITVFVFGASATAMA 396
           : ||:|:|||| :|  :|   :::||   ::     :|  : : |:|| | |  |  :|
Db     334 GIFAAILSTADSQLLVVASTVVKDLYQEVIKKGTKIDEKTALTISRVTVLVVGFLAAILA 393

Qy     397 LLTKTVYGLWYLSSDLVY-IVIF----------PQLLCVLFVKGTNTYGAVAGYVSGLFL 445
                    |:: |:::  |:|            | |:  |: |||  :| :|| : |
Db     394 ----------YVAKDIIFWFVLFAWGGLGASFGPTLILSLYWKGTTKWGVLAGMIVGTIT 443

Qy     446 RITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESG 505
           |      |||:|:                |:| :  |    | : | :  | :| : |
Db     444 TIVW---KLYLKPI----------TGLY-ELVP----AFIFSLIATIIVSMITK------ 479

Qy     506 TLPPK 510
             ||:
Db     480 --PPE 482
```

RESULT 9
AAB85102
ID AAB85102 standard; Protein; 675 AA.
XX
AC AAB85102;
XX
DT 22-AUG-2001 (first entry)
XX
DE Novel human transporter protein (NHP).
XX
KW Novel human protein; NHP; sodium; glucose; co-transporter; therapeutic;
KW diagnostic; pharmacogenomic; gene therapy; mental disorder.
XX
OS Homo sapiens.
XX
PN WO200142469-A1.
XX
PD 14-JUN-2001.
XX
PF 08-DEC-2000; 2000WO-US33363.
XX
PR 10-DEC-1999; 99US-0170137.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
PI Abuin A, Sands AT;
XX
DR WPI; 2001-381689/40.
DR N-PSDB; AAF84022, AAF84023.

XX
PT New nucleic acid molecules encoding human proteins that are
PT structurally similar to sodium-glucose co-transporters, useful for
PT therapeutic, diagnostic, and pharmacogenomic applications -
XX
PS Claim 4; Page 29-30; 34pp; English.
XX
CC The invention relates to a novel human protein (NHP) that is structurally
CC similar to sodium-glucose co-transporters. The NHP protein and DNA are
CC useful in therapeutic, diagnostic and pharmacogenomic applications. An
CC oligopeptide comprising at least 12 amino acids of the NHP is useful for
CC diagnosis, drug screening, clinical trial monitoring and the treatment of
CC physiological disorders. The NHP nucleic acids are useful for identifying
CC mutations associated with a particular disease, and also in a diagnostic
CC or prognostic assay. It is useful for genetically engineering host cells
CC for producing NHP, and in gene therapy. NHP polypeptides, its fragments
CC or mutants are useful for generating antibodies, as reagents in
CC diagnostic assays, for the identification of other cellular gene products
CC related to NHP, and as reagents in assays for screening compounds that
CC are useful as pharmaceutical reagents in the treatment of mental,
CC biological or medical disorders and diseases. The present sequence
CC represents the amino acid sequence of the human NHP.
XX
SQ Sequence 675 AA;

Query Match 10.5%; Score 311.5; DB 22; Length 675;
Best Local Similarity 23.0%; Pred. No. 4.9e-22;
Matches 152; Conservative 112; Mismatches 235; Indels 161; Gaps 28;

```
Qy       2 AFHVEGL----IAIIVFY-LLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGF 56
           ||  :||    ||::| | | :| ||:|:  :||      :  :   : | |:    |
Db      18 AFPQKGLEPGDIAVLVLYFLFVLAVGLWSTVKTK-----RDTVKGYFLAGGDMVWWPVGA 72

Qy      57 TMTATWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLI-LGGLF-------FAKPMRS 108
           :: |: || |:                ||| : |  | |:|   | |||          |:
Db      73 SLFASNVGSGHF-----------IGLAGSSAATGISVSAYELNGLFSVLMLAWIFLPIYI 121

Qy     109 KGYVTMLDPFQQIYGKRMGGLLFIPALMGEMFWAAAIFSAL------GAT-ISVIIDVDM 161
           | || :  : :   || ||:  || ::  ::    ||: :      ||  |   : :|:
Db     122 AGQVTTMPEYLR---KRFGGIR-IPIILAVLYLFIYIFTKISVDMYAGAIFIQQSLHLDL 177

Qy     162 HISVIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVH 221
           :::::    |  :||: ||| :| ||| :|   : :|    :  :  |  ||   |   :
Db     178 YLAIVGLLAITAVYTVAGGLAAVIYTDALQTLIMLIGALTLMGY--SFAAVG--GMEGLK 233

Qy     222 AKY-------------------------------QKPWLGTVDSSEVYS-WLDSFLLLML 249
           ||                                    || | :   : | |
Db     234 EKYFLALASNRSENSSCGLPREDAFHIFRDPLTSDLPWPGVLFGMSIPSLWY-------- 285

Qy     250 GGIPW---QAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYG 306
           |   |   || |:: : ::|:  : :||:  ::     :  | :      :| |
Db     286 ----WCTDQVIVQRTLAAKNLSHAKGGALMAAYLKVLPLFIMVFPGMVSRILFPDQVA-- 339

Qy     307 LPDPKTTEE--------ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSM 358
           ||:  ::         :|:  | ::  | |  :      :  : ||:|||  |   |||::
Db     340 CADPEICQKICSNPSGCSDIAYPKLVLELLPTGLRGLMMAVMVAALMSSLTSIFNSASTI 399

Qy     359 FARNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLW------------ 406
           |  :::    |  ||:||:: | |: |           ||  |  ||
Db     400 FTMDLWN-HLRPRASEKELMIVGRVFV----------LLLVLVSILWIPVVQASQGGQL 447

Qy     407 --YLSSDLVYI-----VIFPQLLCVLFVKGTNTYGAVAGYVSGLFLRITG-GEPYLYLQP 458
           |: |   |:      |:|   :   | | ||  ||  | :||| | :          ::|:||
Db     448 FIYIQSISSYLQPPVAVVF---IMGCFWKRTNEKGAFWGLISGLLLGLVRLVLDFIYVQP 504

Qy     459 LIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDV----- 513
                 ||:  :  :   :    :|: | :| | :| :: :     | ||  ::
Db     505 RC----DQPDERPVLVKSIHYLYFSMILSTVTLITVSTVSWF-----TEPPSKEMVSHLT 555

Qy     514 ----FDAVVARHSEENMDKTILVKNENIKLD-------ELALVKPRQSMTLSSTFTNKEA 562
               | || :          |  ::|   :       :  :|:   | | |   | |::
Db     556 WFTRHDPVVQKEQAPPAAPLSLTLSQNGMPEASSSSSVQFEMVQENTSKTHSCDMTPKQS 615
```

RESULT 10
AAB60093
ID AAB60093 standard; Protein; 675 AA.
XX
AC AAB60093;
XX
DT 28-MAR-2001 (first entry)
XX
DE Human transport protein TPPT-13.
XX
KW Human; transport protein; TPPT; transport disorder; metabolic disorder;
KW neurological disorder; cardiovascular disorder; reproductive disorder;
KW immune disorder; cancer.
XX
OS Homo sapiens.
XX
PN WO200078953-A2.
XX
PD 28-DEC-2000.
XX
PF 16-JUN-2000; 2000WO-US16668.
XX
PR 17-JUN-1999; 99US-0139923.
PR 10-AUG-1999; 99US-0148177.
PR 18-AUG-1999; 99US-0149357.
PR 28-OCT-1999; 99US-0162287.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Lal P, Yang J, Yue H, Hillman JL, Tang YT, Bandman O, Burford N;
PI Baughn MR, Azimzai Y, Lu DAM, Au-Young J, Patterson C;
XX
DR WPI; 2001-041424/05.
DR N-PSDB; AAF27713.
XX
PT Isolated polypeptide with a human transport protein sequence is useful
PT for the diagnosis, prevention and treatment of disorders associated
PT with the immune, reproductive and cardiovascular systems -
XX
PS Claim 2; Page 115-116; 165pp; English.
XX
CC The present invention provides the protein and coding sequences for 43
CC novel human transport proteins (designated TPPTs). These can be used in
CC the diagnosis and treatment of transport, metabolic, neurological,
CC reproductive, cardiovascular and immune disorders, and cell proliferative
CC disorders such as cancer.
XX
SQ Sequence 675 AA;

Query Match 10.4%; Score 310.5; DB 22; Length 675;
Best Local Similarity 23.0%; Pred. No. 6.2e-22;
Matches 151; Conservative 114; Mismatches 237; Indels 155; Gaps 28;

```
Qy       2 AFHVEGL----IAIIVFY-LLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGF 56
           ||  :||    ||::| | | :| ||:|:  :||      :  :   : | |:    |
Db      18 AFPQKGLEPGDIAVLVLYFLFVLAVGLWSTVKTK-----RDTVKGYFLAGGDMVWWPVGA 72

Qy      57 TMTATWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLS-----LILGGLFFAKPMRSKGY 111
           :: |: || |:  | |      | | |   :  | |:      |:|  :|   |:   |
Db      73 SLFASNVGSGHFIGLA------GSGAATGISVSAYELNGLFSVLMLAWIFL--PIYIAGQ 124

Qy     112 VTMLDPFQQIYGKRMGGLLFIPALMGEMFWAAAIFSAL------GAT-ISVIIDVDMHIS 164
           || :  : :    || ||:  || ::  ::    ||: :      ||  |   : :|::::
Db     125 VTTMPEYLR---KRFGGIR-IPIILAVLYLFIYIFTKISVDMYAGAIFIQQSLHLDLYLA 180

Qy     165 VIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY 224
           ::    |  :||: ||| :| ||| :|   : :|    :  :  |  ||   |   : ||
Db     181 IVGLLAITAVYTVAGGLAAVIYTDALQTLIMLIGALTLMGY--SFAAVG--GMEGLKEKY 236

Qy     225 -------------------------------QKPWLGTVDSSEVYS-WLDSFLLLMLGGI 252
                                             || | :   : | |
Db     237 FLALASNRSENSSCGLPREDAFHIFRDPLTSDLPWPGVLFGMSIPSLWY---------- 285
```

```
Qy    253 PW---QAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPD 309
Db    286 -WCTDQVIVQRTLAAKNLSHAKGGALMAAYLKVLPLFIMVFPGMVSRILFPDQVA--CAD 342

Qy    310 PKTTEE--------ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFAR 361
Db    343 PEICQKICSNPSGCSDIAYPKLVLELLPTGLRGLMMAVMVAALMSSLTSIFNSASTIFTM 402

Qy    362 NIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLW----------Y 407
Db    403 DLWN-HLRPRASEKELMIVGRVFV----------LLLVLVSILWIPVVQASQGGQLFIY 450

Qy    408 LSSDLVYI-----VIFPQLLCVLFVKGTNTYGAVAGYVSGLFLRITG-GEPYLYLQPLIF 461
Db    451 IQSISSYLQPPVAVVF---IMGCFWKRTNEKGAFWGLISGLLLGLVRLVLDFIYVQPRC- 506

Qy    462 YPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDV-------- 513
Db    507 ---DQPDERPVLVKSIHYLYFSMILSTVTLITVSTVSWF-----TEPPSKEMVSHLTWFT 558

Qy    514 -FDAVVARHSEENMDKTILVKNENIKLD-------ELALVKPRQSMTLSSTFTNKEA 562
Db    559 RHDPVVQKEQAPPAAPLSLTLSQNGMPEASSSSSVQFEMVQENTSKTHSCDMTPKQS 615
```

RESULT 11
ABB80599
ID ABB80599 standard; Protein; 675 AA.
XX
AC ABB80599;
XX
DT 08-OCT-2002 (first entry)
XX
DE Human sbg1007026SGLT protein.
XX
KW Human; secreted protein; immunosuppressive; cytostatic; nootropic;
KW neuroprotective; antitumour; vulnerary; antimicrobial; ophthalmological;
KW antiparkinsonian; antirheumatic; antiatherosclerotic; dermatological;
KW hypotensive; cerebroprotective; virucide; antiinflammatory; diabetes;
KW malignant tumour; hypertension; hypotension; obesity; bulimia; anorexia;
KW asthma; manic depression; dementia; delirium; mental retardation;
KW Huntington's disease; Tourette's syndrome; schizophrenia;
KW mental disorder; sexual development disorder; blood cascade dysfunction;
KW stroke; growth disorder.
XX
OS Homo sapiens.
XX
PN WO200222802-A1.
XX
PD 21-MAR-2002.
XX
PF 13-SEP-2001; 2001WO-US28462.
XX
PR 13-SEP-2000; 2000US-232455P.
PR 13-SEP-2000; 2000US-232463P.
PR 02-OCT-2000; 2000US-237293P.
PR 07-NOV-2000; 2000US-246269P.
PR 20-NOV-2000; 2000US-252049P.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
PA (GLAX) GLAXO GROUP LTD.
XX
PI Agarwal P, Gogswell JP, Lai Y, Martensen SA, Smith RF, Strum JC;
PI Xie Q;
XX
DR WPI; 2002-393963/42.
DR N-PSDB; ABL57735.
XX
PT Novel isolated secreted polypeptides and polynucleotides encoding them
PT useful for treating cancer, Alzheimer's disease, tumor metastasis,
PT autosomal recessive atypical hemolytic uremic syndrome, wound healing
PT disorder -
XX
PS 1; Page218-219; 246pp; English.
XX
CC The invention relates to a novel isolated polypeptide (ABB80569-ABB80612)
CC (secreted polypeptide) which is encoded by any one of 44 polynucleotide
CC sequences (ABL57705-ABL57748) given in the specification. The
CC polypeptides have immunosuppressive, cytostatic, nootropic,
CC neuroprotective, antitumour, vulnerary, antimicrobial, ophthalmological,
CC antiparkinsonian, antirheumatic, antiatherosclerotic, dermatological,
CC hypotensive, cerebroprotective, virucide, and antiinflammatory activity.
CC The polynucleotide and polypeptide are useful for treating diabetes,
CC malignant tumours, hyper- and hypotension, obesity, bulimia, anorexia,
CC asthma, manic depression, dementia, delirium, mental retardation,
CC Huntington's disease, Tourette's syndrome, schizophrenia, growth, mental
CC or sexual development disorders, and dysfunctions of the blood cascade
CC system including those leading to stroke.
XX
SQ Sequence 675 AA;

Query Match 10.4%; Score 310.5; DB 23; Length 675;
Best Local Similarity 23.0%; Pred. No. 6.2e-22;
Matches 151; Conservative 114; Mismatches 237; Indels 155; Gaps 28;

```
Qy      2 AFHVEGL----IAIIVFY-LLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGF 56
Db     18 AFPQKGLEPGDIAVLVLYFLFVLAVGLWSTVKTK-----RDTVKGYFLAGGDMVWWPVGA 72

Qy     57 TMTATWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLS-----LILGGLFFAKPMRSKGY 111
Db     73 SLFASNVGSGHFIGLA------GSGAATGISVSAYELNGLFSVLMLAWIFL--PIYIAGQ 124

Qy    112 VTMLDPFQQIYGKRMGGLLFIPALMGEMFWAAAIFSAL------GAT-ISVIIDVDMHIS 164
Db    125 VTTMPEYLR---KRFGGIR-IPIILAVLYLFIYIFTKISVDMYAGAIFIQQSLHLDLYLA 180

Qy    165 VIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY 224
Db    181 IVGLLAITAVYTVAGGLAAVIYTDALQTLIMLIGALTLMGY--SFAAVG--GMEGLKEKY 236

Qy    225 ------------------------------QKPWLGTVDSSEVYS-WLDSFLLLMLGGI 252
Db    237 FLALASNRSENSSCGLPREDAFHIFRDPLTSDLPWPGVLFGMSIPSLWY---------- 285

Qy    253 PW---QAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPD 309
Db    286 -WCTDQVIVQRTLAAKNLSHAKGGALMAAYLKVLPLFIMVFPGMVSRILFPDQVA--CAD 342

Qy    310 PKTTEE--------ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFAR 361
Db    343 PEICQKICSNPSGCSDIAYPKLVLELLPTGLRGLMMAVMVAALMSSLTSIFNSASTIFTM 402

Qy    362 NIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLW----------Y 407
Db    403 DLWN-HLRPRASEKELMIVGRVFV----------LLLVLVSILWIPVVQASQGGQLFIY 450

Qy    408 LSSDLVYI-----VIFPQLLCVLFVKGTNTYGAVAGYVSGLFLRITG-GEPYLYLQPLIF 461
Db    451 IQSISSYLQPPVAVVF---IMGCFWKRTNEKGAFWGLISGLLLGLVRLVLDFIYVQPRC- 506

Qy    462 YPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDV-------- 513
Db    507 ---DQPDERPVLVKSIHYLYFSMILSTVTLITVSTVSWF-----TEPPSKEMVSHLTWFT 558

Qy    514 -FDAVVARHSEENMDKTILVKNENIKLD-------ELALVKPRQSMTLSSTFTNKEA 562
Db    559 RHDPVVQKEQAPPAAPLSLTLSQNGMPEASSSSSVQFEMVQENTSKTHSCDMTPKQS 615
```

RESULT 12
AAR73595
ID AAR73595 standard; Protein; 662 AA.
XX

```
AC   AAR73595;
XX
DT   26-DEC-1995  (first entry)
XX
DE   Cotransporter protein SGLT1.
XX
KW   Sodium ion glucose cotransporter; sodium ion nucleoside cotransporter;
KW   SGLT1; SNST1.
XX
OS   Oryctolagus cuniculus.
XX
PN   US5410031-A.
XX
PD   25-APR-1995.
XX
PF   24-FEB-1992;   92US-0841651.
XX
PR   24-FEB-1992;   92US-0841651.
XX
PA   (REGC ) UNIV CALIFORNIA OFFICE TECHNOLOGY.
XX
PI   Pajor AM,  Wright EM;
XX
DR   WPI; 1995-169636/22.
XX
PT   Mammalian sodium ion nucleoside cotransporter protein (SNST) cDNA -
PT   used to produce SNST for identifying nucleoside transport inhibitors
PT   and e.g. nucleoside analogues capable of effective uptake into cells
XX
PS   Disclosure; Columns 29-32; 25pp; English.
XX
CC   cDNA encoding SNST1 was isolated from a rabbit kidney library.  The
CC   encoded protein (AAR73593) showed significant homology to rabbit
CC   intestinal SGLT1.
XX
SQ   Sequence   662 AA;

  Query Match             10.4%;  Score 308.5;  DB 16;  Length 662;
  Best Local Similarity   23.4%;  Pred. No. 9.6e-22;
  Matches  154;  Conservative  110;  Mismatches  238;  Indels  155;  Gaps   26;

Qy      11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING 70
           |:::|::: ||:|| : | | |:         : || :     | :: |: :| |: |
Db      32 IVIYFLVVMAVGLWAMFST-NRGTV----GGFFLAGRSMVWWPIGASLFASNIGSGHFVG 86

Qy      71 TAEAVYVPGYGLAWAQAPIGYS-----LSLILGGLFFAKPMRSKGYVTMLDPFQQIY-GK 124
             |      | | |   |  |:       : ::|| :|    :|: | ||| :  |: : ||
Db      87 LA------GTGAASGIATGGFEWNALIMVVVLGWVFVPIYIRA-GVVTMPEYLQKRFGGK 139

Qy     125 RMGGLLFIPALMGEMFW--AAAIFSALGAT-ISVIIDVDMHISVIISALIATLYTLVGGL 181
           |:   | | :|:  :|   :| |||  ||  | : : :|:::::||  :|  |||: |||
Db     140 RIQIYLSILSLLLYIFTKISADIFS--GAIFIQLTLGLDIYVAIIILLVITGLYTITGGL 197

Qy     182 YSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY----------------Q 225
            :| ||| :|   : ||  |   ||         |: |  ||                  |
Db     198 AAVIYTDTLQTAIMMVGSVILTGFAFHEVG----GYEAFTEKYMRAIPSQISYGNTSIPQ 253

Qy     226 KPWLGTVDSSEVYSWLDSFLLLMLGGIPW------------------QAYFQRVLSSSSA 267
           | :   |:  :      |   : | |||                  |   || ||: :
Db     254 KCYTPREDAFHI------FRDAITGDIPWPGLVFGMSILTLWYWCTDQVIVQRCLSAKNL 307

Qy     268 TYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDP------KTTEEADMILP 321
           :: :    |  :  ::     :::| :      :: |  :|              ::  |
Db     308 SHVKAGCILCGYLKVMPMFLIVMMGMVSRILYTDKVACVVPSECERYCGTRVGCTNIAFP 367

Qy     322 IVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVM 381
            ::  | |  :    |  : |::||| |   |||::|  :||    |: ||:||::
Db     368 TLVVELMPNGLRGLMLSVMMASLMSSLTSIFNSASTLFTMDIY-TKIRKKASEKELMIAG 426

Qy     382 RI-TVFVFGASATAMALLTKTVYG--LWYLSSDLVYI--VIFPQLLCVLFVKGTNTYGAV 436
           |:  :|: | |   : ::     |    |: |   |:    |   | :| |  |  ||
Db     427 RLFMLFLIGISIAWVPIVQSAQSGQLFDYIQSITSYLGPPIAAVFLLAIFWKRVNEPGAF 486

Qy     437 AGYVSGLFLRI----------TG------------GEPYLYLQPLIFYPGYYPDDNGIY 473
             | | |  : |          ||                |  |||   ::|
Db     487 WGLVLGFLIGISRMITEFAYGTGSCMEPSNCPTIICGVHYLYFAIILF------------ 534

Qy     474 NQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAVVA-RHSEENMDKTILV 532
                     | | :| : :|   | :       | : ::    : |:|:|
Db     535 ----------VISIITVVVVSLFTKPI-------PDVHLYRLCWSLRNSKE-------- 568

Qy     533 KNENIKLD--ELALVKPRQSMTLSSTFTNKEAF----------LDVDSSPEGSGTED 577
             | | ||  |  : :  :  | :     |: |           || |  |: :  |:
Db     569 --ERIDLDAGEEDIQEAPEEATDTEVPKKKKGFFRRAYDLFCGLDQDKGPKMTKEEE 623


RESULT 13
ABG30151
ID   ABG30151 standard; Protein; 660 AA.
XX
AC   ABG30151;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #30142.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS94338.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID No 60510; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
```

SQ Sequence 660 AA;

Query Match 10.3%; Score 306; DB 22; Length 660;
Best Local Similarity 22.8%; Pred. No. 1.7e-21;
Matches 148; Conservative 104; Mismatches 218; Indels 178; Gaps 30;

```
Qy      11 IIVFYLLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING 70
           |:::::::: ||:|| : | | |:         : || :     | :: |: :| |:  |
Db      28 IVIYFVVVMAVGLWAMFST-NRGTV----GGFFLAGRSMVWWPIGASLFASNIGSGHFVG 82

Qy      71 TAEAVYVPGYGLAWAQAPIGYS-----LSLILGGLFFAKPMRSK-GYVTMLDPFQQIYGK 124
             |      | | |   |  |:      | ::|| ||   |:  | | |||     :   |
Db      83 LA------GTGAASGIAIGGFEWNALVLVVVLGWLFV--PIYIKAGVVTM----PEYLRK 130

Qy     125 RMGG------------LLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIA 172
           | ||             ||:|   :    :: |||       |:: : ::::::: :   |
Db     131 RFGGQRIQVYLSLLSLLLYIFTKISADIFSGAIF------INLALGLNLYLAIFLLLAIT 184

Qy     173 TLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQK--PWL- 229
           |||: ||| :| ||| :|   : || |  ||         |: |   || |  | :
Db     185 ALYTITGGLAAVIYTDTLQTVIMLVGSLILTGFAFHEVG----GYDAFMEKYMKAIPTIV 240

Qy     230 ---GTVDSSEVYS-WLDSFLLL---MLGGIPW-----------------QAYFQRVLSS 264
               |   : |:   ||| :    : | :||                 |   || ||:
Db     241 SDGNTTFQEKCYTPRADSFHIFRDPLTGDLPWPGFIFGMSILTLWYWCTDQVIVQRCLSA 300

Qy     265 SSATYAQ----VLSFLAAFGCLVMAIPAIL--------IGAI---------GASTDWNQT 303
           : :: :    :  :|        :| :| ::         |  :         |
Db     301 KNMSHVKGGCILCGYLKLMPMFIMVMPGMISRILYTEKIACVVPSECEKYCGTKVGCTNI 360

Qy     304 AYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNI 363
           ||              | ::  | |  :    |  : |::|||  |   |||::|  :|
Db     361 AY--------------PTLVVELMPNGLRGLMLSVMLASLMSSLTSIFNSASTLFTMDI 405

Qy     364 YQLSFRQNASDKEIVWVMRITVFV-FGASATAMALLTKTVYG--LWYLSSDLVYI--VIF 418
           |    |: ||:||::   |: : |  | |   : ::     |    |: |   |:   |
Db     406 Y-AKVRKRASEKELMIAGRLFILVLIGISIAWVPIVQSAQSGQLFDYIQSITSYLGPPIA 464

Qy     419 PQLLCVLFVKGTNTYGAVAGYVSGLFLRI----------TG-------------GEPYLY 455
           |  :| |  |  ||  | : || : |           ||                |  |||
Db     465 AVFLLAIFWKRVNEPGAFWGLILGLLIGISRMITEFAYGTGSCMEPSNCPTIICGVHYLY 524

Qy     456 LQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFD 515
             ::|                          ||:| : || | | :        | : ::
Db     525 FAIILF----------------------AISFITIVVISLLTKPI-------PDVHLYR 554

Qy     516 AV--VARHSEENMDKTILVKNENIKLDELALVKPRQSMTLSSTFTNKE 561
           :    || :|  :  : |||:        |:::: : :    |:
Db     555 LCWSLRNSKEERID--LDAEEENIQ------EGPKETIEIETQVPEKK 594
```

RESULT 14
AAU77134
ID AAU77134 standard; Protein; 675 AA.
XX
AC AAU77134;
XX
DT 05-JUN-2002 (first entry)
XX
DE Human sodium-sugar symporter 32620 polypeptide.
XX
KW Human; sodium-sugar symporter; 32620; neurological disorder; cancer;
KW inflammatory disorder; brain disorder; cellular proliferative disorder;
KW cellular differentiative disorder; cancer; bone metabolism disorder;
KW immune disorder; cardiovascular disorder; viral disease; pain; virucide;
KW metabolic disorder; kidney disorder; intestinal disorder; antidiabetic;
KW antiinflammatory; nootropic; neuroprotective; cytostatic; antirheumatic;
KW antiarthritic; antipsoriatic; antiasthmatic; hypotensive; osteopathic;
KW antiatherosclerotic; antiulcer; analgesic; anorectic; immunomodulator;
KW nephrotropic.
XX
OS Homo sapiens.
XX

PN WO200216582-A2.
XX
PD 28-FEB-2002.
XX
PF 15-AUG-2001; 2001WO-US25475.
XX
PR 22-AUG-2000; 2000US-227068P.
PR 13-AUG-2001; 2001US-0928530.
XX
PA (MILL-) MILLENIUM PHARM INC.
XX
PI Curtis RAJ, Silos-Santiago I;
XX
DR WPI; 2002-280929/32.
DR N-PSDB; ABK10764.
XX
PT Novel isolated human sodium-sugar symporter family member polypeptide,
PT 32620 useful for treating neurological, inflammatory, brain, immune,
PT bone metabolism, kidney and metabolic disorders, pain, viral diseases
PT -
XX
PS Claim 6; Fig 3; 119pp; English.
XX
CC The invention relates to a human sodium-sugar symporter polypeptide
CC termed 32620 and its associated polynucleotide. The sequences of the
CC invention are useful for treating neurological disorders, inflammatory
CC disorders, brain disorders (e.g. Alzheimer's disease), cellular
CC proliferative and/or differentiative disorders (e.g. cancer), disorders
CC associated with bone metabolism (e.g. osteoporosis, rickets), immune
CC disorders (e.g. diabetes mellitus, rheumatoid arthritis, multiple
CC sclerosis, psoriasis, asthma), cardiovascular disorders (e.g.
CC hypertension, atherosclerosis), viral diseases, pain, metabolic disorders
CC (e.g. obesity, cachexia), kidney disorders (e.g. glomerulonephritis) and
CC intestinal disorders (e.g. ulcer). The sequences can also be used to
CC generate antibodies that bind the polypeptide and to locate gene regions
CC associated with genetic disease. This sequence represents the human
CC sodium-sugar symporter 32620 polypeptide.
XX
SQ Sequence 675 AA;

Query Match 10.0%; Score 298.5; DB 23; Length 675;
Best Local Similarity 22.7%; Pred. No. 1e-20;
Matches 149; Conservative 115; Mismatches 238; Indels 155; Gaps 29;

```
Qy       2 AFHVEGL----IAIIVFY-LLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGF 56
           ||  :||    ||::| | | :| ||:|:  :||     :       :  |  :   ||
Db      18 AFPQKGLEPGDIAVLVLYFLFVLAVGLWSTVKTKR----DTVKGYFLAEGNMVWWPVGA- 72

Qy      57 TMTATWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLS-----LILGGLFFAKPMRSKGY 111
           :: |: || |:  | |     . | | |    :   | |:     |:|  :|   |:   |
Db      73 SLFASNVGSGHFIGLA-----.-GSGAATGISVSAYELNGLFSVLMLAWIFL--PIYIAGQ 124

Qy     112 VTMLDPFQQIYGKRMGGLLFIPALMGEMFWAAAIFSALGATI---SVIID----VDMHIS 164
           || :   : :   || ||:  || ::  ::    ||: :   :   :: |      :|::::
Db     125 VTTMPEYLR---KRFGGIR-IPIILAVLYLFIYIFTKISVDMYAGAIFIQQSSHLDLYLA 180

Qy     165 VIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY 224
           ::    |  :||: ||| :| ||| :|   : :|    : :  |  ||    |   :  ||
Db     181 IVGLLAITAVYTVAGGLAAVIYTDALQTLIMLIGALTLMGY--SFAAVG--GMEGLKEKY 236

Qy     225 ------------------------------QKPWLGTVDSSEVYS-WLDSFLLLMLGGI 252
                                           || | :    : | |
Db     237 FLALASNRSENSSCGLPREDAFHIFRDPLTSDLPWPGVLFGMSIPSLWY---------- 285

Qy     253 PW---QAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPD 309
           |    |   || |:: : ::|:  : :||:  ::      :  | :      :| |    |
Db     286 -WCTDQVIVQRTLAAKNLSHAKGGALMAAYLKVLPLFIMVFPGMVSRILFPDQVA--CAD 342

Qy     310 PKTTEE--------ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFAR 361
           |:  ::         :|:  | ::  | |  :     :  : ||:|||  |   |||::|
Db     343 PEICQKICSNPSGCSDIAYPKLVLELLPTGLRGLMMAVMVAALMSSLTSIFNSASTIFTM 402

Qy     362 NIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLW----------Y 407
```

```
          :::    |  ||:||::  | |: |           ||   |  ||                |
Db    403 DLWN-HLRPRASEKELMIVGRVFV----------LLLVLVSILWIPVVQASQGGQLFIY 450

Qy    408 LSSDLVYI-----VIFPQLLCVLFVKGTNTYGAVAGYVSGLFLRITG-GEPYLYLQPLIF 461
          : |   |:     |:|   :   | | ||  ||  | :||| | :      ::|:||
Db    451 IQSISSYLQPPVAVVF---IMGCFWKRTNEKGAFWGLISGLLLGLVRLVLDFIYVQPRC- 506

Qy    462 YPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDV-------- 513
              ||:  :  :   :  :|: | :| | :| :: :     | ||  ::
Db    507 ---DQPDERPVLVKSIHYLYFSMILSTVTLITVSTVSWF-----TEPPSKEMVSHLTWFT 558

Qy    514 -FDAVVARHSEENMDKTILVKNENIKLD-------ELALVKPRQSMTLSSTFTNKEA 562
           | || :          |  ::|   :       :  :|:   | | |   | |::
Db    559 RHDPVVQKEQAPPAAPLSLTLSQNGMPEASSSSSVQFEMVQENTSKTHSCDMTPKQS 615


RESULT 15
AAR73593
ID   AAR73593 standard; Protein; 672 AA.
XX
AC   AAR73593;
XX
DT   26-DEC-1995  (first entry)
XX
DE   Cotransporter protein SNST1.
XX
KW   Sodium ion nucleoside cotransporter protein; SNST1; rabbit.
XX
OS   Oryctolagus cuniculus.
XX
PN   US5410031-A.
XX
PD   25-APR-1995.
XX
PF   24-FEB-1992;   92US-0841651.
XX
PR   24-FEB-1992;   92US-0841651.
XX
PA   (REGC ) UNIV CALIFORNIA OFFICE TECHNOLOGY.
XX
PI   Pajor AM,  Wright EM;
XX
DR   WPI; 1995-169636/22.
DR   N-PSDB; AAQ89779.
XX
PT   Mammalian sodium ion nucleoside cotransporter protein (SNST) cDNA -
PT   used to produce SNST for identifying nucleoside transport inhibitors
PT   and e.g. nucleoside analogues capable of effective uptake into cells
XX
PS   Disclosure; Columns 17-22; 25pp; English.
XX
CC   cDNAs were isolated from a rabbit kidney library by high stringency
CC   hybridization with a rabbit sodium/glucose cotransporter (SGLT1)
CC   cDNA.  Clone RK-C was demonstrated by expression studies to encode
CC   most of an SNST protein.  A composite sequence (AAQ89779) was obtd. by
CC   combining RK-C and another clone, RK-44.  The cDNA was used to
CC   express SNST1 in Xenopus oocytes.
XX
SQ   Sequence   672 AA;

  Query Match             10.0%;  Score 298;  DB 16;  Length 672;
  Best Local Similarity   25.0%;  Pred. No. 1.1e-20;
  Matches  153;  Conservative   89;  Mismatches  232;  Indels  138;  Gaps   25;

Qy      9 IAII-VFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY 67
          ||:| .::||:: ||:|:  || | |:         : || :     | :: |: :| |:
Db     26 IAVIAAYFLLVIGVGLWSMCRT-NRGTV----GGYFLAGRSMVWWPVGASLFASNIGSGH 80

Qy     68 INGTAEAVYVPGYGLAWAQAPIGYSLS----LILGGLFFAKPMRSKGYVTMLDPFQQIYG 123
           | |      | | |   |  |:  :    ::| |  ||    : | :||     |
Db     81 FVGLA------GTGAANGLAVAGFEWNALFVVLLLGWLFAPVYLTAGVITM----PQYLR 130

Qy    124 KRMGG-----------LLFIPALMGEMFWAAAIF--SALGATISVIIDVDMHISVIISA 169
```

```
          || ||              |:|   :    :: |:|   |||  |        : |||
Db    131 KRFGGHRIRLYLSVLSLFLYIFTKISVDMFSGAVFIQQALGWNI--------YASVIALL 182

Qy    170 LIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY----- 224
           |  :||: ||| :: ||| || | |  | :|   :|          |:: :  ||
Db    183 GITMVYTVTGGLAALMYTDTVQTFVIIAGAFILTGYAFHEVG----GYSGLFDKYMGAMT 238

Qy    225 -----QKPWLGTVDSSEVYSWLDSFLLL---MLGGIPW-----------------QAYF 258
              : | :| : ||      ||: ||   : | :||                    |
Db    239 SLTVSEDPAVGNISSSCYRPRPDSYHLLRDPVTGDLPWPALLLGLTIVSGWYWCSDQVIV 298

Qy    259 QRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKT----TE 314
          || |:  : |: :    |  :  |      :: | |      :: |  |:       ||
Db    299 QRCLAGRNLTHIKAGCILCGYLKLTPMFLMVMPGMISRILYPDEVACVAPEVCKRVCGTE 358

Qy    315 E--ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNA 372
             :::  | ::  | |  :    |  : ||:|||  |   |:|::|  :|| |  |  |
Db    359 VGCSNIAYPRLVVKLMPNGLRGLMLAVMLAALMSSLRSIFNSSSTLFTMDIYTL--RPRA 416

Qy    373 SDKEIVWVMRITVFVFGASATAMALLTKTVYG---LWYLSSDLVYIV--IFPQLLCVLFV 427
           : |:: | |: |    | : |   : :   |     |: |   |:   :    :  |||
Db    417 GEGELLLVGRLWVVFIVAVSVAWLPVVQAAQGGQLFDYIQSVSSYLAPPVSAVFVVALFV 476

Qy    428 KGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMV-- 485
             |  ||  | : || : :   |                              :| | | : |
Db    477 PRVNEKGAFWGLIGGLLMGLARLIP----------------------EFSFGTGSCVRP 513

Qy    486 ---TSFLTNICISYLAKYLFE-SG---------TLP-PKLDVFDAVVA-RHSEENMDKTI 530
              :||  :   | |  ||  ||          | | |:  :   | : |||:|
Db    514 SACPAFLCRVHYLYFAIVLFFCSGLLIIIVSLCTAPIPRKHLHRLVFSLRHSKE------ 567

Qy    531 LVKNENIKLDEL 542
            : |::  |||
Db    568 --EREDLDADEL 577
```

Search completed: March 31, 2003, 18:34:31
Job time : 70 secs

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: March 31, 2003, 18:30:51 ; Search time 64 Seconds
(without alignments)
1867.302 Million cell updates/sec

Title: US-10-069-541-6
Perfect score: 2972
Sequence: 1 MAFHVEGLIAIIVFYLLILL.....EAFLDVDSSPEGSGTEDNLQ 580

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2972	100.0	580	4	Q9GZV3	Q9gzv3 homo sapien
2	2820	94.9	580	11	Q9JMD7	Q9jmd7 rattus norv
3	2795	94.0	580	11	Q99PK3	Q99pk3 mus musculu
4	2791	93.9	580	11	Q9ESW5	Q9esw5 mus musculu
5	2253	75.8	584	13	Q8UWF0	Q8uwf0 torpedo mar
6	1557.5	52.4	614	5	Q961W3	Q961w3 drosophila
7	1530.5	51.5	579	5	Q9GPB1	Q9gpb1 limulus pol
8	1453	48.9	576	5	O02228	O02228 caenorhabdi
9	723.5	24.3	278	5	Q9VE46	Q9ve46 drosophila
10	381.5	12.8	479	16	Q8Y273	Q8y273 ralstonia s
11	344	11.6	492	17	Q9V2P3	Q9v2p3 pyrococcus
12	316	10.6	493	17	Q8U3M8	Q8u3m8 pyrococcus
13	314.5	10.6	665	11	Q9QXI6	Q9qxi6 mus musculu
14	312.5	10.5	665	11	Q9QXX5	Q9qxx5 mus musculu
15	310.5	10.4	675	4	Q8WWX8	Q8wwx8 homo sapien
16	309	10.4	670	11	Q923I7	Q923i7 mus musculu
17	307.5	10.3	675	4	Q96PP5	Q96pp5 homo sapien
18	301	10.1	463	16	Q9I3S6	Q9i3s6 pseudomonas
19	300	10.1	500	16	Q9CN55	Q9cn55 pasteurella
20	299.5	10.1	507	16	Q9K9E2	Q9k9e2 bacillus ha
21	299	10.1	662	6	Q9BDF6	Q9bdf6 equus cabal
22	292	9.8	698	4	Q8WY15	Q8wy15 homo sapien
23	291	9.8	514	17	Q8TUR0	Q8tur0 methanosarc
24	287	9.7	678	11	Q8VDT1	Q8vdt1 mus musculu
25	286.5	9.6	674	6	Q28728	Q28728 oryctolagus
26	286	9.6	491	17	O58753	O58753 pyrococcus
27	284	9.6	537	17	O59086	O59086 pyrococcus
28	283	9.5	484	17	Q9V083	Q9v083 pyrococcus
29	282.5	9.5	501	16	Q9Z9W5	Q9z9w5 bacillus ha
30	281.5	9.5	673	13	O42135	O42135 xenopus lae
31	279.5	9.4	496	16	Q9KKN1	Q9kkn1 vibrio chol
32	277	9.3	597	6	Q28610	Q28610 oryctolagus
33	274.5	9.2	639	17	Q8TTZ6	Q8ttz6 methanosarc
34	272.5	9.2	524	2	O32355	O32355 corynebacte
35	272.5	9.2	616	11	Q9ET36	Q9et36 mus musculu
36	272.5	9.2	718	6	O19135	O19135 bos taurus
37	272	9.2	660	11	Q91ZP4	Q91zp4 mus musculu
38	271	9.1	537	17	Q9V0K5	Q9v0k5 pyrococcus
39	268	9.0	490	2	Q93N71	Q93n71 bacillus ce
40	267	9.0	522	17	Q9HR16	Q9hr16 halobacteri
41	266.5	9.0	484	16	Q8RFL0	Q8rfl0 fusobacteri
42	263.5	8.9	494	2	Q57023	Q57023 pseudomonas
43	263.5	8.9	504	16	Q9CNT4	Q9cnt4 pasteurella
44	262.5	8.8	477	16	Q9CLW1	Q9clw1 pasteurella
45	262.5	8.8	497	2	Q8VQN4	Q8vqn4 vibrio vuln

ALIGNMENTS

RESULT 1
Q9GZV3
ID Q9GZV3 PRELIMINARY; PRT; 580 AA.
AC Q9GZV3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE High affinity choline transporter (High-affinity choline transporter
DE CHT1).
GN CHT1 OR SLC5A7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HYPOTHALAMUS;
RA Bruess M.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=HYPOTHALAMUS;
RA Wieland A., Bonisch H., Bruss M.;
RT "Molecular cloning of the human and murine high affinity choline
RT transporters and characterization ofthe human gene-structure.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20483599; PubMed=11027560;
RA Apparsundaram S., Ferguson S.M., George A.L. Jr., Blakely R.D.;
RT "Molecular cloning of a human, hemicholinium-3-sensitive choline
RT transporter.";
RL Biochem. Biophys. Res. Commun. 276:862-867(2000).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=SPINAL CORD;
RX PubMed=11068039;
RA Okuda T., Haga T.;
RT "Functional characterization of the human high-affinity choline

```
RT   transporter.";
RL   FEBS Lett. 484:92-97(2000).
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Bruess M.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   Wieland A., Bonisch H., Bruss M.;
RT   "Molecular cloning of the human and murine high affinity choline
RT   transporters and characterizationof the human gene structure.";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ401466; CAC03717.1; -.
DR   EMBL; AF276871; AAG25940.1; -.
DR   EMBL; AB043997; BAB18161.1; -.
DR   EMBL; AJ308378; CAC88115.1; -.
DR   EMBL; AJ308379; CAC88115.1; JOINED.
DR   EMBL; AJ308380; CAC88115.1; JOINED.
DR   EMBL; AJ308381; CAC88115.1; JOINED.
DR   EMBL; AJ308382; CAC88115.1; JOINED.
DR   EMBL; AJ308383; CAC88115.1; JOINED.
DR   EMBL; AJ308384; CAC88115.1; JOINED.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
SQ   SEQUENCE   580 AA;  63203 MW;  66CB35496CB6E2D6 CRC64;

  Query Match           100.0%;  Score 2972;  DB 4;  Length 580;
  Best Local Similarity  100.0%;  Pred. No. 4.3e-205;
  Matches  580;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60

Qy      61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120

Qy     121 IYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 IYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180

Qy     181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSW 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSW 240

Qy     241 LDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDW 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 LDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDW 300

Qy     301 NQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 NQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360

Qy     361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420

Qy     421 LLCVLFVKGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 LLCVLFVKGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480

Qy     481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAVVARHSEENMDKTILVKNENIKLD 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAVVARHSEENMDKTILVKNENIKLD 540

Qy     541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
          ||||||||||||||||||||||||||||||||||||||||||
Db     541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580

RESULT 2
Q9JMD7
ID   Q9JMD7      PRELIMINARY;      PRT;   580 AA.
AC   Q9JMD7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   High-affinity choline transporter CHT1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WISTAR; TISSUE=SPINAL CORD;
RX   MEDLINE=20116099; PubMed=10649566;
RA   Okuda T., Haga T., Kanai Y., Endou H., Ishihara T., Katsura I.;
RT   "Identification and characterization of the high-affinity choline
RT   transporter.";
RL   Nat. Neurosci. 3:120-125(2000).
DR   EMBL; AB030947; BAA90484.1; -.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
SQ   SEQUENCE   580 AA;  63406 MW;  B7CB73323DAD17A7 CRC64;

  Query Match           94.9%;  Score 2820;  DB 11;  Length 580;
  Best Local Similarity  93.1%;  Pred. No. 3.4e-194;
  Matches  540;  Conservative   24;  Mismatches   16;  Indels    0;  Gaps    0;

Qy       1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
          | ||||||:|||:|||||| ||||||||:|||||:||||||||||||||||||||||||||
Db       1 MPFHVEGLVAIILFYLLIFLVGIWAAWKTKNSGNAEERSEAIIVGGRDIGLLVGGFTMTA 60

Qy      61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
          |||||||||||||| || ||||||||||||||||||||||||||||||||||||||||||||
Db      61 TWVGGGYINGTAEAVYGPGCGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120

Qy     121 IYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180
          ||||||||||||||||||||||||||||||||||||||::||||:||||| |||||||
Db     121 IYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDVNISVIVSALIAILYTLVGG 180

Qy     181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSW 240
          ||||||||||||||:|||||||||||||| |||||||||| |||||::| |||:|
Db     181 LYSVAYTDVVQLFCIFIGLWISVPFALSHPAVTDIGFTAVHAKYQSPWLGTIESVEVYTW 240

Qy     241 LDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDW 300
          ||:|||||||||||||||||||||||||||||||||||||||||:||| ||||||||||
Db     241 LDNFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMALPAICIGAIGASTDW 300

Qy     301 NQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360
          |||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 NQTAYGFPDPKTKEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360

Qy     361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||
Db     361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIIIFPQ 420

Qy     421 LLCVLFVKGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480
          ||||||:|||||||||||: |||||||||||||||||||||||| ||||||:||||
Db     421 LLCVLFIKGTNTYGAVAGYIFGLFLRITGGEPYLYLQPLIFYPGYYPDKNGIYNQRFPFK 480

Qy     481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAVVARHSEENMDKTILVKNENIKLD 540
          ||:||||| ||||:|||||||||||||||:|||||:|||||||||||||:|||||||:
Db     481 TLSMVTSFFTNICVSYLAKYLFESGTLPPKLDIFDAVVSRHSEENMDKTILVRNENIKLN 540

Qy     541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
          ||| ||||||:|||||||||| ||||||||||||||||||
Db     541 ELAPVKPRQSLTLSSTFTNKEALLDVDSSPEGSGTEDNLQ 580

RESULT 3
Q99PK3
ID   Q99PK3      PRELIMINARY;      PRT;   580 AA.
```

```
AC   Q99PK3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Sodium and chloride-dependent high-affinity choline transporter.
GN   SLC5A7.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Apparsundaram S., Ferguson S.M., Blakely R.D.;
RT   "Molecular cloning and characterization of human and murine high-
RT   affinity choline transporters.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF276872; AAG36945.2; -.
DR   MGD; MGI:1927126; Slc5a7.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
SQ   SEQUENCE   580 AA;  63383 MW;  DDBF58ED428270AF CRC64;

  Query Match             94.0%;  Score 2795;  DB 11;  Length 580;
  Best Local Similarity   92.6%;  Pred. No. 2.1e-192;
  Matches  537;  Conservative   23;  Mismatches   20;  Indels    0;  Gaps    0;

Qy       1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
Db       1 MPFHVEGLVAIILFYLLIFLVGIWAAWKTKNSGNPEERSEAIIVGGRDIGLLVGGFTMTA 60

Qy      61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
Db      61 TWVGGGYINGTAEAVYGPGCGLAWAHAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFKQ 120

Qy     121 IYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180
Db     121 IYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDVNISVIVSALIAILYTLVGG 180

Qy     181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSW 240
Db     181 LYSVAYTDVVQLFCIFIGLWISVPFALSHPAVTDIGFTAVHAKYQSPWLGTIESVEVYTW 240

Qy     241 LDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDW 300
Db     241 LDNFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMALPAICIGAIGASTDW 300

Qy     301 NQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360
Db     301 NQTAYGYPDPKTKEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360

Qy     361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420
Db     361 RNIYQLSFRQNASDKEIVWVMRITVLVFGASATAMALLTKTVYGLWYLSSDLVYIIIFPQ 420

Qy     421 LLCVLFVKGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480
Db     421 LLCVLFIKGTNTYGAVAGYIFGLFLRITGGEPYLYLQPLIFYPGYYSDKNGIYNQRFPFK 480

Qy     481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAVVARHSEENMDKTILVKNENIKLD 540
Db     481 TLSMVTSFFTNICVSYLAKYLFESGTLPPKLDVFDAVVARHSEENMDKTILVRNENIKLN 540

Qy     541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
Db     541 ELAPVKPRQSLTLSSTFTNKEALLDVDSSPEGSGTEDNLQ 580


RESULT 4
Q9ESW5
ID   Q9ESW5      PRELIMINARY;      PRT;   580 AA.
AC   Q9ESW5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   High affinity choline transporter.
GN   SLC5A7 OR CHT1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/CJ; TISSUE=BRAIN STEM;
RA   Bruess M.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/CJ; TISSUE=BRAIN STEM;
RA   Wieland A., Bonisch H., Bruss M.;
RT   "Molecular cloning of the human and murine high affinity choline
RT   transporters and characterization ofthe human gene-structure.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ401467; CAC03719.1; -.
DR   MGD; MGI:1927126; Slc5a7.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
SQ   SEQUENCE   580 AA;  63331 MW;  A4F1387CAA9EAAFE CRC64;

  Query Match             93.9%;  Score 2791;  DB 11;  Length 580;
  Best Local Similarity   92.4%;  Pred. No. 4.2e-192;
  Matches  536;  Conservative   24;  Mismatches   20;  Indels    0;  Gaps    0;

Qy       1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
Db       1 MSFHVEGLVAIILFYLLIFLVGIWAAWKTKNSGNPEEHSEAIIVGGRDIGLLVGGFTMTA 60

Qy      61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
Db      61 TWVGGGYINGTAVAVYGPGCGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120

Qy     121 IYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180
Db     121 IYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDVNISVIVSALIAILYTLVGG 180

Qy     181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSW 240
Db     181 LYSVAYTDVVQLFCIFIGLWISVPFALSHPAVTDIGFTAVHAKYQSPWLGTIESVEVYTW 240

Qy     241 LDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDW 300
Db     241 LDNFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSYLAAFGCLVMALPAICIGAIGASTDW 300

Qy     301 NQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360
Db     301 NQTAYGYPDPKTKEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360

Qy     361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420
Db     361 RNIYQLSFRQNASDKEIVWVMRITVLVFGASATAMALLTKTVYGLWYLSSDLVYIIIFPQ 420

Qy     421 LLCVLFVKGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480
Db     421 LLCVLFIKGTNTYGAVAGYIFGLFLRITGGEPYLYLQPLIFYPGYYSDKNGIYNQRFPFK 480

Qy     481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAVVARHSEENMDKTILVKNENIKLD 540
Db     481 TLSMVTSFFTNICVSYLAKYLFESGTLPPKLDVFDAVVARHSEENMDKTILVRNENIKLN 540

Qy     541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
Db     541 ELAPVKPRQSLTLSSTFTNKEALLDVDSSPEGSGTEDNLQ 580


RESULT 5
```

```
Q8UWF0
ID   Q8UWF0      PRELIMINARY;      PRT;   584 AA.
AC   Q8UWF0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   High affinity choline transporter.
GN   CHT1.
OS   Torpedo marmorata (Marbled electric ray).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Squalea; Hypnosqualea; Pristiorajea; Batoidea;
OC   Torpediniformes; Torpedinoidei; Torpedinidae; Torpedo.
OX   NCBI_TaxID=7788;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=ELECTRIC LOBE;
RA   Guermonprez L., O'Regan S., Meunier F.M., Morot-Gaudry-Talarmain Y.;
RT   "Cyclosporin, FK506 and rapamycin inhibit neuronal choline uptake via
RT   calcineurin-dependent and independent mechanisms.";
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ420808; CAD12727.1; -.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
SQ   SEQUENCE   584 AA;  63660 MW;  995F937B01195A3D CRC64;

  Query Match           75.8%;  Score 2253;  DB 13;  Length 584;
  Best Local Similarity   72.4%;  Pred. No. 1.6e-153;
  Matches  418;  Conservative   83;  Mismatches   70;  Indels    6;  Gaps    2;

Qy      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSG--SAEERSEAIIVGGRDIGLLVGGFTM 58
          |  |::|::||::|||||| ||:||||::||:    | :||||::|||||||||||||
Db      1 MTVHIDGIVAIVLFYLLILFVGLWAAWKSKNTSMEGAMDRSEAIMIGGRDIGLLVGGFTM 60

Qy     59 TATWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPF 118
          ||||||||||||||||||||||||||||||| ||:|||::||||||||||:|||||||||
Db     61 TATWVGGGYINGTAEAVYVPGYGLAWAQAPFGYALSLVIGGLFFAKPMRSRGYVTMLDPF 120

Qy    119 QQIYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLV 178
          ||:|||||||||||||:||:||:||| ||||||:|||:|:::::||::||:|| |||||
Db    121 QQMYGKRMGGLLFIPALLGEIFWSAAILSALGATLSVIVDININVSVVVSAVIAVLYTLV 180

Qy    179 GGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVY 238
          |||||||||||||||||:|||||:|||| :||| ||  ||    ||:||:| : | :
Db    181 GGLYSVAYTDVVQLFCIFLGLWISIPFALLNPAVTDIIVTANQEVYQEPWVGNIQSKDSL 240

Qy    239 SWLDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGAST 298
          |:|:|||||||||||| |||||||:||||||||||||||||::||||:|||||| ||
Db    241 IWIDNFLLLMLGGIPWQVYFQRVLSASSATYAQVLSFLAAFGCVLMAIPSVLIGAIGTST 300

Qy    299 DWNQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSM 358
          |||||:|||| |    | |||||||||:||| ||||||||||||||||||||||||||||
Db    301 DWNQTSYGLPGPIGKNETDMILPIVLQHLCPPYISFFGLGAVSAAVMSSADSSILSASSM 360

Qy    359 FARNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIF 418
          |||||| |:||| |||||||||||||:|:|| :||:|||| :::|||||||||||||:||
Db    361 FARNIYHLAFRQEASDKEIVWVMRITIFLFGGAATSMALLAQSIYGLWYLSSDLVYVIIF 420

Qy    419 PQLLCVLFVKGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPD----DNGIYN 474
          |||: |||||||||||::|||: |  |||:|||||::|| |:||| | |    |: :|
Db    421 PQLISVLFVKGTNTYGSIAGYIIGFLLRISGGEPYLHMQPFIYYPGCYLDHSFGDDPVYV 480

Qy    475 QKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAVVARHSEENMDKTILVKN 534
          |:|||||:||: ||| |  :||| |||| || ||||||  |:||::||:| |||| |:
Db    481 QRFPFKTMAMLFSFLGNTGVSYLVKYLFVSGILPPKLDFLDSVVSKHSKEIMDKTFLMNQ 540

Qy    535 ENIKLDELALVKPRQSMTLSSTFTNKEAFLDVDSSPE 571
          :|| | ||  | |  | ::|:  |||||| |:: :||
Db    541 DNITLSELVHVNPIHSASVSAALTNKEAFEDIEPNPE 577


RESULT 6
Q961W3
```

```
ID   Q961W3      PRELIMINARY;      PRT;   614 AA.
AC   Q961W3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   GH02984p.
GN   CG7708.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY047521; AAK77253.1; -.
DR   FlyBase; FBgn0038641; CG7708.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
SQ   SEQUENCE   614 AA;  66893 MW;  71A77E1216360042 CRC64;

  Query Match           52.4%;  Score 1557.5;  DB 5;  Length 614;
  Best Local Similarity   58.5%;  Pred. No. 1.4e-103;
  Matches  302;  Conservative   84;  Mismatches  115;  Indels   15;  Gaps    6;

Qy      4 HVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWV 63
          :: |:::|::|||||:|||||  | | ||:  |  | ::: || ||| || ||||||||
Db      3 NIAGVVSIVLFYLLILVVGIWAG-RKKQSGNDSE--EEVMLAGRSIGLFVGIFTMTATWV 59

Qy     64 GGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQIYG 123
          ||||||||||:|    || | ||| ||:|||: ||:||| ||| :||:||||| |  :|
Db     60 GGGYINGTAEAIYTS--GLVWCQAPFGYALSLVFGGIFFANPMRKQGYITMLDPLQDSFG 117

Qy    124 KRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGGLYS 183
          :|||||||:||| ||:|||| | :|||||:|||||:|   |||:|: ||  ||| |||||
Db    118 ERMGGLLFLPALCGEVFWAAGILAALGATLSVIIDMDHRTSVILSSCIAIFYTLFGGLYS 177

Qy    184 VAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSWLDS 243
          |||||||:||||||:|||: :||| |:  |  :     :  :  |:| |:  : : ::|
Db    178 VAYTDVIQLFCIFIGLWMCIPFAWSNEHVGSL------SDLEVDWIGHVEPKKHWLYIDY 231

Qy    244 FLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQT 303
          |||: |||||| |||||||| :|  ||:||::|| ||::|||| :|||||  :| ||:|
Db    232 GLLLVFGGIPWQVYFQRVLSSKTAGRAQLLSYVAAAGCILMAIPPVLIGAIAKATPWNET 291

Qy    304 AYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNI 363
          |  | | | :|  ||||:||||| | ::|||||||||||||||||:|||:||||||:
Db    292 DYKGPYPLTVDETSMILPMVLQYLTPDFVSFFGLGAVSAAVMSSADSSVLSAASMFARNV 351

Qy    364 YQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQLLC 423
          |:| ||| ||: ||:|||| : | |  || |||   ::|||| : |||||:::|||||
Db    352 YKLIFRQKASEMEIIWVMRVAIIVVGILATIMALTIPSIYGLWSMCSDLVYVILFPQLLM 411

Qy    424 VL-FVKGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTL 482
          |: | |  ||||::: |:  | :|::|||  | | ||| |||   |    | |||:|:
Db    412 VVHFKKHCNTYGSLSAYIVALAIRLSGGEAILGLAPLIKYPGY---DEETKEQMFPFRTM 468

Qy    483 AMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAVV 518
          ||: | :| | :|:  | :|||| |||  | |  ||
Db    469 AMLLSLVTLISVSWWTKMMFESGKLPPSYDYFRCVV 504


RESULT 7
Q9GPB1
ID   Q9GPB1      PRELIMINARY;      PRT;   579 AA.
AC   Q9GPB1;
```

```
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Choline cotransporter.
OS   Limulus polyphemus (Atlantic horseshoe crab).
OC   Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;
OC   Limulidae; Limulus.
OX   NCBI_TaxID=6850;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21261948; PubMed=11368908;
RA   Wang Y., Cao Z., Newkirk R.F., Ivy M.T., Townsel J.G.;
RT   "Molecular cloning of a cDNA for a putative choline co-transporter
RT   from Limulus CNS.";
RL   Gene 268:123-131(2001).
DR   EMBL; AY011119; AAG41055.1; -.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
SQ   SEQUENCE   579 AA;  62937 MW;  FE7F29D4FAF47F04 CRC64;

  Query Match             51.5%;  Score 1530.5;  DB 5;  Length 579;
  Best Local Similarity   52.0%;  Pred. No. 1.1e-101;
  Matches  305;  Conservative  115;  Mismatches  134;  Indels   33;  Gaps   11;

Qy       1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
Db       1 MAVNILGVVSIGIFYVIILIVGIWAS-RKKKTSSGQSETEEIMLAGRNIGFLVGVLTMTA 59

Qy      61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
Db      60 TWVGGGYINGTAEAMY--NNGLVWCQAPFGYALSLFIGGIVFAKKMRSQGYVTMLDPLQE 117

Qy     121 IYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180
Db     118 NFGSKMGGLLFLPALCGEIFWSAAILAALGATISVITELESSTSIIVSSSIAVFYTFFGG 177

Qy     181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSW 240
Db     178 FYSVAYTDVIQLFCIFFGLWLCIPFSFSHEAVGSLS--------SIDFLGSVKLSDAGIN 229

Qy     241 LDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDW 300
Db     230 VDIWLLLIFGGIPWQVYFQRVLSAKNVSNAQVLSYVAAVGCVVMAIPAILIGVIAKATAW 289

Qy     301 NQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360
Db     290 NETALGM--PLTPNDTSLVLPLVLHYLTPTAVSFFGLGAVSAAVMSSSDSSILSASSLFS 347

Qy     361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420
Db     348 RNVYKLIFRQKASEREVVWVIRISILVVGILATAMALTVKSVYGLWYLSSDLIYVILFPQ 407

Qy     421 LLCVLFVKG-TNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPF 479
Db     408 LLCVVHLKKYCNTYGSLSAYIVGFLLRALGGESILGLEPVIHYP-FFSETSG---QRFPF 463

Qy     480 KTLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAVVARHSEENMDKTILVKNENIK- 538
Db     464 RTLSMLASLITLLAISGITKWIFEMNHLPAKLDIFRCVT--NIQEN-----IIKIQKLQG 516

Qy     539 -----LDEL--ALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDN 578
Db     517 GAMPVLDSIKKEIYQKDMNNSFNTVVNSGNAELLTDSTYSGKIKKNN 563


RESULT 8
O02228
ID   O02228      PRELIMINARY;      PRT;   576 AA.
AC   O02228; Q9NL58;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   C48D1.3 protein (High-affinity choline transporter CHO-1).
GN   C48D1.3 OR CHO-1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Burton J.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=N2;
RX   MEDLINE=20116099; PubMed=10649566;
RA   Okuda T., Haga T., Kanai Y., Endou H., Ishihara T., Katsura I.;
RT   "Identification and characterization of the high-affinity choline
RT   transporter.";
RL   Nat. Neurosci. 3:120-125(2000).
DR   EMBL; Z81049; CAB02847.2; -.
DR   EMBL; AB030946; BAA90483.1; -.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
SQ   SEQUENCE   576 AA;  62427 MW;  FAB09778358288D9 CRC64;

  Query Match             48.9%;  Score 1453;  DB 5;  Length 576;
  Best Local Similarity   50.5%;  Pred. No. 3.9e-96;
  Matches  295;  Conservative   95;  Mismatches  150;  Indels   44;  Gaps    9;

Qy       7 GLIAIIVFYLLILLVGIWAAWRTKNSGSAEER----SEAIIVGGRDIGLLVGGFTMTATW 62
Db       6 GIVAIVFFYVLILVVGIWAGRKSKSSKELESEAGAATEEVMLAGRNIGTLVGIFTMTATW 65

Qy      63 VGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQIY 122
Db      66 VGGAYINGTAEALY--NGGLLGCQAPVGYAISLVMGGLLFAKKMREEGYITMLDPFQHKY 123

Qy     123 GKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGGLY 182
Db     124 GQRIGGLMYVPALLGETFWTAAILSALGATLSVILGIDMNASVTLSACIAVFYTFTGGYY 183

Qy     183 SVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDS-SEVYSWL 241
Db     184 AVAYTDVVQLFCIFVGLWVCVPAAMVHDGAKDISRNA------GDWIGEIGGFKETSLWI 237

Qy     242 DSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWN 301
Db     238 DCMLLLVFGGIPWQVYFQRVLSSKTAHGAQTLSFVAGVGCILMAIPPALIGAIARNTDWR 297

Qy     302 QTAYGLPDPKTTEEA------DMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSA 355
Db     298 MTDYSPWNNGTKVESIPPDKRNMVVPLVFQYLTPRWVAFIGLGAVSAAVMSSADSSVLSA 357

Qy     356 SSMFARNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYI 415
Db     358 ASMFAHNIWKLTIRPHASEKEVIIVMRIAIICVGIMATIMALTIQSIYGLWYLCADLVYV 417

Qy     416 VIFPQLLCVLFVKGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQ 475
Db     418 ILFPQLLCVVYMPRSNTYGSLAGYAVGLVLRLIGGEPLVSLPAFFHYPMY---TDGV--Q 472

Qy     476 KFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAVV---ARHSEENMDKTILV 532
Db     473 YFPFRTTAMLSSMATIYIVSIQSEKLFKSGRLSPEWDVMGCVVNIPIDHVPLPSDVSFAV 532

Qy     533 KNENIKL----------------DELALVKPRQSMTLSSTFTN 559
```

```
          :| : :                    ||  |: |    :  || :|
Db    533 SSETLNMKAPNGTPAPVHPNQQPSDENTLLHPYSDQSYYSTNSN 576


RESULT 9
Q9VE46
ID   Q9VE46      PRELIMINARY;      PRT;   278 AA.
AC   Q9VE46;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   CG7708 protein.
GN   CG7708.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003723; AAF55583.1; -.
DR   FlyBase; FBgn0038641; CG7708.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
SQ   SEQUENCE   278 AA;  30105 MW;  04A1DB76071B739B CRC64;

  Query Match           24.3%;  Score 723.5;  DB 5;  Length 278;
  Best Local Similarity  64.2%;  Pred. No. 3.5e-44;
  Matches  136;  Conservative   33;  Mismatches   38;  Indels    5;  Gaps    3;

Qy       4 HVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWV 63
           :: |:::|::|||||||:|||||  | | ||:  |  | ::: || ||| || ||||||||

Db     3 NIAGVVSIVLFYLLILVVGIWAG-RKKQSGNDSE--EEVMLAGRSIGLFVGIFTMTATWV 59

Qy      64 GGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQIYG 123
           |||||||||||:|    || | ||| ||:|||: ||:||| ||| :||:||||| |  :|
Db     60 GGGYINGTAEAIYTS--GLVWCQAPFGYALSLVFGGIFFANPMRKQGYITMLDPLQDSFG 117

Qy     124 KRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGGLYS 183
           :|||||||:||| ||:|||| | :|||||:|||||:|   |||:|: ||  ||| |||||
Db     118 ERMGGLLFLPALCGEVFWAAGILAALGATLSVIIDMDHRTSVILSSCIAIFYTLFGGLYS 177

Qy     184 VAYTDVVQLFCIFVGLWISVPFALSHPAVADI 215
           ||||||:|||||||:|||: :||| |:  |  :
Db     178 VAYTDVIQLFCIFIGLWMCIPFAWSNEHVGSL 209


RESULT 10
Q8Y273
ID   Q8Y273      PRELIMINARY;      PRT;   479 AA.
AC   Q8Y273;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Probable sodium/solute symporter transmembrane protein.
GN   RSC0463 OR RS04434.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OC   Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC   Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
DR   EMBL; AL646059; CAD13991.1; -.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
KW   Complete proteome.
SQ   SEQUENCE   479 AA;  52091 MW;  560962E411DBC9B8 CRC64;

  Query Match           12.8%;  Score 381.5;  DB 16;  Length 479;
  Best Local Similarity  28.2%;  Pred. No. 2.3e-19;
  Matches  128;  Conservative   85;  Mismatches  188;  Indels   53;  Gaps   16;

Qy      11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING 70
           :||:::: : :|:||| | :|:           | || :   |   |: ||| |   : |
Db       6 VIVYWVISVGIGLWAALRVRNTAD-------FAVAGRGLPFYVVTATVFATWFGSETVLG 58

Qy      71 TAEAVYVPGYGLAWAQA-PIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQIYGKRMGGL 129
             ||::   ||    | | | || ||| |||||:|:     :|: | ::  :|:    |
Db      59 -IPAVFLK-EGLHGVVADPFGSSLCLILVGLFFARPLYRMNLLTIGDFYRNRFGRVAEVL 116

Qy     130 LFIPALMGEMFWAAAIFSALGATISVIID--VDMHISVIISALIATLYTLVGGLYSVAYT 187
           :  ::  : | ||   |||     : |  :     ::| |    :||| ||::||| |
Db     117 TTLCIVVSYLGWVAAQIKALGLVFYTVSDGGLSQQTGMMIGAASVLVYTLFGGMWSVAVT 176

Qy     188 DVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHA----KYQKPWLGTVDSSEVYSWLDS 243
           | :|:  | :|: : : : :|  |   :     ||    |:  :    :  ||  :: :
Db     177 DFIQMIIIVIGM-MYIGWEVSGQA-GGVATVVAHASAAGKFS--FWPAFNPIEVIGFVTA 232

Qy     244 FLLLMLGGIPWQAYFQRVLSS-------SSATYAQVLSFLAAFGCLVMAIPAILIGAIGA 296
           :: :||| || |  |||| ||       :::    || || ||  : :|  | ||
Db     233 WITMMLGSIPQQDVFQRVTSSRTERIAGTASVLGGVLYFLFAFIPMFLAYSATLI----- 287

Qy     297 STDWNQTAYGLPDPK-----TTEEADMILP-IVLQYLCPVYISFFGLGAVSAAVMSSADS 350
                     ||:       :: :||| :||::  |:       ||: :|: | | :
```

```
Db    288 ------------DPQMVARYINTDSQLILPKLVLEH-APLVAQVMFFGALLSAIKSCASA 334

Qy    351 SILSASSMFARNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTK-TVYGLWYLS 409

Db    335 TLLAPSVTFAENVLR-PMLPRMDDKRFLRVMQAVVLVFTALVTLFALNSHLSIFHMVENA 393

Qy    410 SDLVYIVIFPQLLCVLFVKGTNTYGAVAGYVSGL 443

Db    394 YKVTLVAAFVPLAFGLFWKRATRQGGLLAIALGL 427


RESULT 11
Q9V2P3
ID   Q9V2P3      PRELIMINARY;      PRT;   492 AA.
AC   Q9V2P3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Proline symporter (Proline permease).
GN   PUTP-3 OR PAB2354.
OS   Pyrococcus abyssi.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=29292;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ORSAY;
RA   Heilig R.;
RT   "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT   structure and evolution.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ248283; CAB48955.1; -.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   TIGRFAMs; TIGR00813; sss; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
KW   Complete proteome.
SQ   SEQUENCE   492 AA;  53457 MW;  A7C72B1AF29282B3 CRC64;

  Query Match             11.6%;  Score 344;  DB 17;  Length 492;
  Best Local Similarity   24.2%;  Pred. No. 1.2e-16;
  Matches  132;  Conservative   99;  Mismatches  196;  Indels  118;  Gaps   25;

Qy      8 LIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY 67

Db     14 LVAFLFTLILPILVGFYAMKRTKS-------EEDFFVGGRAMDKITVALSAVSSGRSSWL 66

Qy     68 INGTAEAVYVPGYGLAWAQAPIGYSLS----LILGGLFFAKPMRSKGYVTMLDPFQQIYG 123

Db     67 VLGLSGMAYKMGVTAVW--AAVGYIVAEMFQFVYMGIRLRKFSERFNAITVPDYFEARFR 124

Qy    124 K-----RMGG----LLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATL 174

Db    125 DTSKILRIAASIIIIIFLTSYVGAQFNAGA------KTLSTALGISIFTALMISVLMIIV 178

Qy    175 YTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFT--------AVHAKYQK 226

Db    179 YMILGGFIAVAYNDVIRAVIMIIGLVV-----LPVIAVAKVGGTEEVLKVLHALDPKLIN 233

Qy    227 PW---LGTVDSSEVYSWLDSFLLLMLG-GIPWQAY-FQRVLSSSSATYAQVLSFLAAFGC 281

Db    234 PWAFGAGVVIG-----------FLGIGFGSPGQPHIIVRYMSIDDPNKLRVSTVVGTFWN 282

Qy    282 LVMAIPAILIGAIGASTDWNQTAYGLPDPKTT--EEADMILP-IVLQYLCPVYISFFGLG 338

Db    283 VVLAWGAIFVGLAGRAI--------VPDVSQLPGKNAEMIYPYLSAQYFPPILYGIL-IG 333

Qy    339 AVSAAVMSSADSSILSASSMFARNIYQLSFRQNA--SDKEIVWVMRITVFVFGASATAMA 396

Db    334 GIFAAILSTADSQLLVVASTVVKDLYQEVIKKGTKIDEKTALTISRVTVLVVGFLAAILA 393

Qy    397 LLTKTVYGLWYLSSDLVY-IVIF----------PQLLCVLFVKGTNTYGAVAGYVSGLFL 445

Db    394 ----------YVAKDIIFWFVLFAWGGLGASFGPTLILSLYWKGTTKWGVLAGMIVGTIT 443

Qy    446 RITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESG 505

Db    444 TIVW---KLYLKPI----------TGLY-ELVP----AFIFSLIATIIVSMITK------ 479

Qy    506 TLPPK 510

Db    480 --PPE 482


RESULT 12
Q8U3M8
ID   Q8U3M8      PRELIMINARY;      PRT;   493 AA.
AC   Q8U3M8;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative proline permease.
GN   PF0429.
OS   Pyrococcus furiosus.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=2261;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA   Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT   "The complete sequence of the Pyrococcus furiosus genome.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE010165; AAL80553.1; -.
KW   Complete proteome; Hypothetical protein.
SQ   SEQUENCE   493 AA;  53431 MW;  46D94B3C6175ADDD CRC64;

  Query Match             10.6%;  Score 316;  DB 17;  Length 493;
  Best Local Similarity   24.4%;  Pred. No. 1.2e-14;
  Matches  132;  Conservative   95;  Mismatches  203;  Indels  112;  Gaps   23;

Qy      8 LIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY 67

Db     14 LAAFLFTLLLPILVGFLVMRRTKT-------EEDFFVGGRAMDKIVVALSAVSSGRSSWL 66

Qy     68 INGTAEAVYVPGYGLAWAQAPIGYSLS----LILGGLFFAKPMRSKGYVTMLDPFQQIYG 123

Db     67 VLGLSGMAYARGVSAVW--AAVGYIVAEMFQFVYMGIRLRKFSEKYNCITVPDYFEERFR 124

Qy    124 KRMGGLLFIPALMGEMF---WAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180

Db    125 DDSKIVRLVASIIIAIFLTAYVGAQFNAGAKSLSTALGLSTLTALLIAALMIIIYMVLGG 184

Qy    181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIG--------FTAVHAKYQKPW-LGT 231

Db    185 FIAVAYNDVIRAVIMIFGLTV-----LPAVAVAKIGGVNVMLSILKALDPKYIDPLSLGV 239

Qy    232 VDSSEVYSWLDSFLLLMLGGIPWQAY-FQRVLSSSSATYAQVLSFLAAFGCLVMAIPAIL 290

Db    240 -------GGLIGFLGIGLGS-PGQPHIIVRYMAIDDPDRLRASTVIGTFWNVVLAWGAIF 291

Qy    291 IGAIGAS--TDWNQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFG-------LGAVS 341

Db    292 VGLAGRALYPDVNM----LPD----ANSEMIYLVLSS-------DFFGPLLYGLLIGGVF 336

Qy    342 AAVMSSADSSILSASSMFARNIYQ--LSFRQNASDKEIVWVMRITVFVFGASATAMALLT 399

Db    337 AAILSTADSQLLVVASTIVRDFYQEIIKKGEKLSEEKAVFLSRIVVFLAGIFAMLLA--- 393

Qy    400 KTVYGLWYLSSDLVY-IVIF----------PQLLCVLFVKGTNTYGAVAGYVSGLFLRIT 448

Db    394 -------YFAKDIIFWLVLFAWGGLGAAIGPTLIVSLYWKRATKWGVVAGMIVGALTTII 446

Qy    449 GGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLP 508

Db    447 W---KLYIRPI----------TGLYE---------LVPAFL----FALIATVLVSLATKP 480
```

```
Qy     509 PK 510
           |:
Db     481 PE 482


RESULT 13
Q9QXI6
ID   Q9QXI6      PRELIMINARY;      PRT;   665 AA.
AC   Q9QXI6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   SGLT1 protein (Similar to solute carrier family 5, member alpha
DE   1).
GN   SLC5A1 OR SGLT1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Tabatabai N.M., Blumenthal S.S., Petering D.H.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF208031; AAF17249.1; -.
DR   EMBL; BC003845; AAH03845.1; -.
DR   MGD; MGI:107678; Slc5a1.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   TIGRFAMs; TIGR00813; sss; 1.
DR   PROSITE; PS00456; NA_SOLUT_SYMP_1; 1.
DR   PROSITE; PS00457; NA_SOLUT_SYMP_2; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
SQ   SEQUENCE   665 AA;  73317 MW;  71AC8F7C488D9D59 CRC64;

  Query Match             10.6%;  Score 314.5;  DB 11;  Length 665;
  Best Local Similarity   24.3%;  Pred. No. 2.2e-14;
  Matches  139;  Conservative   90;  Mismatches  206;  Indels  137;  Gaps   23;

Qy      11 IIVFYLLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY--- 67
           |:::::::: ||:|| : | | |:         : || :     | :: |: :| |:
Db      32 IVIYFVVVMAVGLWAMFST-NRGTV----GGFFLAGRSMVWWPIGASLFASNIGSGHFVG 86

Qy      68 INGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSK-GYVTMLDPFQQIYGKRM 126
           : || | :  | |        | ::|| :|   |:  | | |||     :   ||
Db      87 LAGTGAAAGIAMGGFEWN----ALVLVVVLGWIFV--PIYIKAGVVTM----PEYLRKRF 136

Qy     127 GG------------LLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATL 174
           ||            ||:|  :    :: |||      |:: : :|:::: |   | |
Db     137 GGKRIQIYLSVLSLLLYIFTKISADIFSGAIF------INLALGLDIYLAIFILLAITAL 190

Qy     175 YTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDS 234
           ||: ||| :| ||| :|   : || :|   || :       |: |   || |     | :
Db     191 YTITGGLAAVIYTDTLQTAIMLVGSFILTGFAFNEVG----GYEAFMDKYMKAIPTKVSN 246

Qy     235 ------SEVYS-WLDSFLLL---MLGGIPW-----------------QAYFQRVLSSSS 266
                | |:   ||| :    : | :||                   |   || ||: :
Db     247 GNFTAKEECYTPRADSFHIFRDPITGDMPWPGLIFGLAILALWYWCTDQVIVQRCLSAKN 306

Qy     267 ATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKTTEE----------- 315
           :: :   | :  |:     :: | |      : |  ||     ||
Db     307 MSHVKAGCTLCGYLKLLPMFLMVMPGMISRILYTEKIACVLP-----EECQKYCGTPVGC 361

Qy     316 ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDK 375
           ::  | ::  | |  :    |  : |::||| |   |||::|  :||    |: ||:|
Db     362 TNIAYPTLVVELMPNGLRGLMLSVMMASLMSSLTSIFNSASTLFTMDIY-TKIRKKASEK 420

Qy     376 EIVWVMRITVFV-FGASATAMALLTKTVYG--LWYLSSDLVYI--VIFPQLLCVLFVKGT 430
           |::   |: : |  | |   : ::     |    |: |   |:   |   :| |
```

```
Db     421 ELMIAGRLFILVLIGISIAWVPIVQSAQSGQLFDYIQSITSYLGPPIAAVFLLAIFCKRV 480

Qy     431 NTYGAVAGYVSGLFLRI----------TG------------GEPYLYLQPLIFYPGYYP 467
           |  ||  | : |  : |          ||             |  |||   ::|
Db     481 NEQGAFWGLILGFLIGISRMITEFAYGTGSCMEPSNCPKIICGVHYLYFAIILF------ 534

Qy     468 DDNGIYNQKFPFKTLAMVTSFLTNICISYLAK 499
                           | | :| : ||:| |
Db     535 -----------------VISVITILIISFLTK 549


RESULT 14
Q9QXX5
ID   Q9QXX5      PRELIMINARY;      PRT;   665 AA.
AC   Q9QXX5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Sodium-glucose cotransporter 1.
GN   SLC5A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yang Q., Tian Y., Wada J., Kashihara N., Wallner E.I., Kanwar Y.S.;
RT   "Expression characteristics and relevance of mouse sodium glucose
RT   cotransporter (SGLT-1) in embryonic renal development.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF163846; AAF23269.1; -.
DR   MGD; MGI:107678; Slc5a1.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   TIGRFAMs; TIGR00813; sss; 1.
DR   PROSITE; PS00456; NA_SOLUT_SYMP_1; 1.
DR   PROSITE; PS00457; NA_SOLUT_SYMP_2; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
SQ   SEQUENCE   665 AA;  73109 MW;  E4DA3585A2098C23 CRC64;

  Query Match             10.5%;  Score 312.5;  DB 11;  Length 665;
  Best Local Similarity   23.6%;  Pred. No. 3.1e-14;
  Matches  138;  Conservative   90;  Mismatches  195;  Indels  161;  Gaps   25;

Qy      11 IIVFYLLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY--- 67
           | ::::::: ||:|| : | | |:         : || :     | :: |: :| |:
Db      32 IAIYFVVVMAVGLWAMFST-NRGTV----GGFFLAGRSMVWWPIGASLFASNIGSGHFVG 86

Qy      68 INGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSK-GYVTMLDPFQQIYGKRM 126
           : || | :  | |        | ::|| :|   |:  | | |||     :   ||
Db      87 LAGTGAAAGIAMGGFEWN----ALVLVVVLGWIFV--PIYIKAGVVTM----PEYLRKRF 136

Qy     127 GG------------LLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATL 174
           ||            ||:|  :    :: |||      |:: : :|:::: |   | |
Db     137 GGKRIQIYLSVLSLLLYIFTKISADIFSGAIF------INLTLGLDIYLAIFILLAITAL 190

Qy     175 YTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDS 234
           ||: ||| :| ||| :|   : || :|   || :       |: |   || |     | :
Db     191 YTITGGLAAVIYTDTLQTAIMLVGSFILTGFAFNEVG----GYEAFMDKYMKAIPTKVSN 246

Qy     235 ------SEVYS-WLDSFLLL---MLGGIPW-----------------QAYFQRVLSSSS 266
                | |:   ||| :    : | :||                   |   || ||: :
Db     247 GNLTAKEECYTPRADSFHIFRDPITGDMPWPGLIFGLAILALWYWCTDQVIVQRCLSAKN 306

Qy     267 ATYAQ----VLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKTTEEADMILPI 322
           :: :    :  :|      :| :| ::          ::  |       ||:   :||
Db     307 MSHVKAGCTLCGYLKLLPMFLMVMPGMI----------SRILY-------TEKIACVLPE 349

Qy     323 VLQYLC------------------PVYISFFGLGAVSAAVMSSADSSILSASSMFARNI 363
           | |                    |  :    |  : |::||| |   |||::|  :|
Db     350 ECQKYCGTPVGCPTSATPTLVVELMPNGLQGLKLSGMMASLMSSLTSIFNSASTLFTMDI 409

Qy     364 YQLSFRQNASDKEIVWVMRITVFV-FGASATAMALLTKTVYG--LWYLSSDLVYI--VIF 418
```

```
Db    410 Y-TKIRKKASEKELMIAGRLFILVLIGISIAWVPIVQSAQSGQLFDYIQSITSYLGPPIA 468

Qy    419 PQLLCVLFVKGTNTYGAVAGYVSGLFLRI----------TG------------GEPYLY 455

Db    469 AVFLLAIFCKRVNEQGAFWGLILGFLIGISRMITEFAYGTGSCMEPSNCPKIICGVHYLY 528

Qy    456 LQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAK 499

Db    529 FAIILF----------VISVITILIISFLTK 549


RESULT 15
Q8WWX8
ID   Q8WWX8      PRELIMINARY;      PRT;   675 AA.
AC   Q8WWX8;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Sodium/glucose cotransporter KST1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Roll P., Massacrier A., Pereira S., Robaglia-Schlupp A., Cau P.,
RA   Szepetowski P.;
RT   "New human sodium/glucose cotransporter gene (hKST1).";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY044906; AAK97784.1; -.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   TIGRFAMs; TIGR00813; sss; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
SQ   SEQUENCE   675 AA;  74036 MW;  44F7BA6D2FE92335 CRC64;

  Query Match             10.4%;  Score 310.5;  DB 4;  Length 675;
  Best Local Similarity   23.0%;  Pred. No. 4.3e-14;
  Matches  151;  Conservative  114;  Mismatches  237;  Indels  155;  Gaps   28;

Qy      2 AFHVEGL----IAIIVFY-LLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGF 56

Db     18 AFPQKGLEPGDIAVLVLYFLFVLAVGLWSTVKTK-----RDTVKGYFLAGGDMVWWPVGA 72

Qy     57 TMTATWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLS-----LILGGLFFAKPMRSKGY 111

Db     73 SLFASNVGSGHFIGLA------GSGAATGISVSAYELNGLFSVLMLAWIFL--PIYIAGQ 124

Qy    112 VTMLDPFQQIYGKRMGGLLFIPALMGEMFWAAAIFSAL------GAT-ISVIIDVDMHIS 164

Db    125 VTTMPEYLR---KRFGGIR-IPIILAVLYLFIYIFTKISVDMYAGAIFIQQSLHLDLYLA 180

Qy    165 VIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY 224

Db    181 IVGLLAITAVYTVAGGLAAVIYTDALQTLIMLIGALTLMGY--SFAAVG--GMEGLKEKY 236

Qy    225 ------------------------------QKPWLGTVDSSEVYS-WLDSFLLLMLGGI 252

Db    237 FLALASNRSENSSCGLPREDAFHIFRDPLTSDLPWPGVLFGMSIPSLWY----------- 285

Qy    253 PW---QAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPD 309

Db    286 -WCTDQVIVQRTLAAKNLSHAKGGALMAAYLKVLPLFIMVFPGMVSRILFPDQVA--CAD 342

Qy    310 PKTTEE--------ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFAR 361

Db    343 PEICQKICSNPSGCSDIAYPKLVLELLPTGLRGLMMAVMVAALMSSLTSIFNSASTIFTM 402

Qy    362 NIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLW----------Y 407

Db    403 DLWN-HLRPRASEKELMIVGRVFV----------LLLVLVSILWIPVVQASQGGQLFIY 450

Qy    408 LSSDLVYI-----VIFPQLLCVLFVKGTNTYGAVAGYVSGLFLRITG-GEPYLYLQPLIF 461

Db    451 IQSISSYLQPPVAVVF---IMGCFWKRTNEKGAFWGLISGLLLGLVRLVLDFIYVQPRC- 506

Qy    462 YPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDV------- 513

Db    507 ---DQPDERPVLVKSIHYLYFSMILSTVTLITVSTVSWF-----TEPPSKEMVSHLTWFT 558

Qy    514 -FDAVVARHSEENMDKTILVKNENIKLD-------ELALVKPRQSMTLSSTFTNKEA 562

Db    559 RHDPVVQKEQAPPAAPLSLTLSQNGMPEASSSSSVQFEMVQENTSKTHSCDMTPKQS 615


Search completed: March 31, 2003, 18:37:43
Job time : 69 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: March 31, 2003, 17:49:41 ; Search time 38 Seconds
(without alignments)
633.060 Million cell updates/sec

Title: US-10-069-541-6
Perfect score: 2972
Sequence: 1 MAFHVEGLIAIIVFYLLILL.....EAFLDVDSSPEGSGTEDNLQ 580

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	308.5	10.4	662	1	SL51_RABIT	P11170 oryctolagus
2	308	10.4	670	1	SL52_RAT	P53792 rattus norv
3	306	10.3	664	1	SL51_HUMAN	P13866 homo sapien
4	306	10.3	665	1	SL51_RAT	P53790 rattus norv
5	303.5	10.2	660	1	SL54_PIG	P31636 sus scrofa
6	298	10.0	672	1	SL52_RABIT	P26430 oryctolagus
7	294	9.9	659	1	SL54_HUMAN	Q9ny91 homo sapien
8	294	9.9	664	1	SL51_SHEEP	P53791 ovis aries
9	293.5	9.9	543	1	SGLT_VIBPA	P96169 vibrio para
10	292	9.8	672	1	SL52_HUMAN	P31639 homo sapien
11	290	9.8	656	1	SL54_MOUSE	Q9et37 mus musculu
12	289	9.7	718	1	SL53_MOUSE	Q9jkz2 mus musculu
13	285	9.6	492	1	OPUE_BACSU	O06493 bacillus su
14	278.5	9.4	718	1	SL53_CANFA	P31637 canis famil
15	275	9.3	718	1	SL53_BOVIN	P53793 bos taurus
16	272.5	9.2	718	1	SL53_HUMAN	P53794 homo sapien
17	260.5	8.8	504	1	PUTP_HAEIN	P45174 haemophilus
18	260.5	8.8	605	1	SL51_PIG	P26429 sus scrofa
19	253	8.5	502	1	PUTP_ECOLI	P07117 escherichia
20	249	8.4	502	1	PUTP_SALTY	P10502 salmonella
21	239.5	8.1	484	1	PANF_HAEIN	P44963 haemophilus
22	239	8.0	549	1	YJCG_ECOLI	P32705 escherichia
23	231	7.8	643	1	SL55_HUMAN	Q92911 homo sapien
24	229.5	7.7	636	1	SL56_RABIT	Q9xt77 oryctolagus
25	226	7.6	571	1	YIDK_ECOLI	P31448 escherichia
26	219.5	7.4	634	1	SL56_RAT	O70247 rattus norv
27	218	7.3	618	1	SL55_RAT	Q63008 rattus norv
28	209	7.0	635	1	SL56_HUMAN	Q9y289 homo sapien
29	206.5	6.9	483	1	PANF_ECOLI	P16256 escherichia
30	198.5	6.7	735	1	DUR3_YEAST	P33413 saccharomyc
31	198	6.7	516	1	YWCA_BACSU	P39599 bacillus su
32	137	4.6	657	1	NU5M_EMENI	P11628 emericella
33	132	4.4	482	1	ARCD_PSEAE	P18275 pseudomonas
34	132	4.4	492	1	YCLF_BACSU	P94408 bacillus su
35	132	4.4	499	1	MVIN_AQUAE	O67658 aquifex aeo
36	131	4.4	570	1	NU5M_PARTE	P15584 paramecium
37	125.5	4.2	499	1	NQOE_PARDE	P29926 paracoccus
38	125	4.2	476	1	YAAJ_ECOLI	P30143 escherichia
39	124.5	4.2	666	1	NU5M_CHOCR	P48920 chondrus cr
40	123	4.1	517	1	QOX1_SULAC	P98004 sulfolobus
41	121	4.1	692	1	NU5C_MARPO	P06264 marchantia
42	120	4.0	687	1	CSTA_HELPJ	Q9zk47 helicobacte
43	120	4.0	741	1	YBIO_ECOLI	P75783 escherichia
44	119	4.0	468	1	YQKI_BACSU	P54571 bacillus su
45	119	4.0	599	1	THIY_YEAST	Q08579 saccharomyc

ALIGNMENTS

```
RESULT 1
SL51_RABIT
ID   SL51_RABIT     STANDARD;      PRT;   662 AA.
AC   P11170;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Sodium/glucose cotransporter 1 (Na(+)/glucose cotransporter 1)
DE   (High affinity sodium-glucose cotransporter).
GN   SLC5A1 OR SGLT1.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=New Zealand white;
RX   MEDLINE=88065856; PubMed=2446136;
RA   Hediger M.A., Coady M.J., Ikeda T.S., Wright E.M.;
RT   "Expression cloning and cDNA sequencing of the Na+/glucose co-
RT   transporter.";
RL   Nature 330:379-381(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=New Zealand white; TISSUE=Kidney cortex;
RX   MEDLINE=91223090; PubMed=2025641;
RA   Morrison A.I., Panayotova-Heiermann M., Feigl G., Schoelermann B.,
RA   Kinne R.K.H.;
RT   "Sequence comparison of the sodium-D-glucose cotransport systems in
RT   rabbit renal and intestinal epithelia.";
RL   Biochim. Biophys. Acta 1089:121-123(1991).
CC   -!- FUNCTION: ACTIVELY TRANSPORTS GLUCOSE INTO CELLS BY NA+
CC       CO-TRANSPORT WITH A NA+ TO GLUCOSE COUPLING RATIO OF 2:1.
CC   -!- FUNCTION: EFFICIENT SUBSTRATE TRANSPORT IN MAMMALIAN KIDNEY IS
CC       PROVIDED BY THE CONCERTED ACTION OF A LOW AFFINITY HIGH CAPACITY
CC       AND A HIGH AFFINITY LOW CAPACITY NA(+)/GLUCOSE COTRANSPORTER
CC       ARRANGED IN SERIES ALONG KIDNEY PROXIMAL TUBULES.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: FOUND PREDOMINANTLY IN INTESTINE AND IN OUTER
CC       RENAL MEDULLA.
CC   -!- DISEASE: MUTATION OF ASP-28 IS IMPLICATED IN GLUCOSE/GALACTOSE
CC       MALABSORPTION.
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X06419; CAA29727.1; -.
DR   EMBL; X55355; CAA39040.1; -.
DR   PIR; S00515; S00515.
DR   PIR; S15974; S15974.
```

```
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   TIGRFAMs; TIGR00813; sss; 1.
DR   PROSITE; PS00456; NA_SOLUT_SYMP_1; 1.
DR   PROSITE; PS00457; NA_SOLUT_SYMP_2; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
KW   Transport; Sugar transport; Transmembrane; Sodium transport; Symport;
KW   Glycoprotein.
FT   DOMAIN        1     28       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     29     47       POTENTIAL.
FT   DOMAIN       48     64       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     65     85       POTENTIAL.
FT   DOMAIN       86    105       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    106    126       POTENTIAL.
FT   DOMAIN      127    171       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    172    191       POTENTIAL.
FT   DOMAIN      192    208       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    209    229       POTENTIAL.
FT   DOMAIN      230    270       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    271    291       POTENTIAL.
FT   DOMAIN      292    314       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    315    334       POTENTIAL.
FT   DOMAIN      335    423       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    424    443       POTENTIAL.
FT   DOMAIN      444    455       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    456    476       POTENTIAL.
FT   DOMAIN      477    526       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    527    547       POTENTIAL.
FT   DOMAIN      548    640       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    641    661       POTENTIAL.
FT   DOMAIN      662    662       EXTRACELLULAR (POTENTIAL).
FT   CARBOHYD    248    248       N-LINKED (GLCNAC. . .).
FT   SITE         43     43       IMPLICATED IN SODIUM COUPLING
FT                                (BY SIMILARITY).
FT   SITE        300    300       IMPLICATED IN SODIUM COUPLING
FT                                (BY SIMILARITY).
SQ   SEQUENCE   662 AA;  73079 MW;  03F55A0309CBBE01 CRC64;

  Query Match              10.4%;  Score 308.5;  DB 1;  Length 662;
  Best Local Similarity    23.4%;  Pred. No. 5e-13;
  Matches  154;  Conservative  110;  Mismatches  238;  Indels  155;  Gaps   26;

Qy     11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING 70
Db     32 IVIYFLVVMAVGLWAMFST-NRGTV----GGFFLAGRSMVWWPIGASLFASNIGSGHFVG 86

Qy     71 TAEAVYVPGYGLAWAQAPIGYS-----LSLILGGLFFAKPMRSKGYVTMLDPFQQIY-GK 124
Db     87 LA------GTGAASGIATGGFEWNALIMVVVLGWVFVPIYIRA-GVVTMPEYLQKRFGGK 139

Qy    125 RMGGLLFIPALMGEMFW--AAAIFSALGAT-ISVIIDVDMHISVIISALIATLYTLVGGL 181
Db    140 RIQIYLSILSLLLYIFTKISADIFS--GAIFIQLTLGLDIYVAIIILLVITGLYTITGGL 197

Qy    182 YSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY----------------Q 225
Db    198 AAVIYTDTLQTAIMMVGSVILTGFAFHEVG----GYEAFTEKYMRAIPSQISYGNTSIPQ 253

Qy    226 KPWLGTVDSSEVYSWLDSFLLLMLGGIPW----------------QAYFQRVLSSSSA 267
Db    254 KCYTPREDAFHI------FRDAITGDIPWPGLVFGMSILTLWYWCTDQVIVQRCLSAKNL 307

Qy    268 TYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDP------KTTEEADMILP 321
Db    308 SHVKAGCILCGYLKVMPMFLIVMMGMVSRILYTDKVACVVPSECERYCGTRVGCTNIAFP 367

Qy    322 IVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVM 381
Db    368 TLVVELMPNGLRGLMLSVMMASLMSSLTSIFNSASTLFTMDIY-TKIRKKASEKELMIAG 426

Qy    382 RI-TVFVFGASATAMALLTKTVYG--LWYLSSDLVYI--VIFPQLLCVLFVKGTNTYGAV 436
Db    427 RLFMLFLIGISIAWVPIVQSAQSGQLFDYIQSITSYLGPPIAAVFLLAIFWKRVNEPGAF 486

Qy    437 AGYVSGLFLRI----------TG------------GEPYLYLQPLIFYPGYYPDDNGIY 473
Db    487 WGLVLGFLIGISRMITEFAYGTGSCMEPSNCPTIICGVHYLYFAIILF------------ 534

Qy    474 NQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAVVA-RHSEENMDKTILV 532
Db    535 -----------VISIITVVVVSLFTKPI-------PDVHLYRLCWSLRNSKE-------- 568

Qy    533 KNENIKLD--ELALVKPRQSMTLSSTFTNKEAF----------LDVDSSPEGSGTED 577
Db    569 --ERIDLDAGEEDIQEAPEEATDTEVPKKKKGFFRRAYDLFCGLDQDKGPKMTKEEE 623


RESULT 2
SL52_RAT
ID   SL52_RAT       STANDARD;      PRT;   670 AA.
AC   P53792;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Sodium/glucose cotransporter 2 (Na(+)/glucose cotransporter 2)
DE   (Low affinity sodium-glucose cotransporter).
GN   SLC5A2 OR SGLT2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Kidney;
RX   MEDLINE=96094332; PubMed=7493971;
RA   You G., Lee W.-S., Barros E.J.G., Kanai Y., Huo T.-L., Khawaja S.,
RA   Wells R.G., Nigam S.K., Hediger M.A.;
RT   "Molecular characteristics of Na(+)-coupled glucose transporters in
RT   adult and embryonic rat kidney.";
RL   J. Biol. Chem. 270:29365-29371(1995).
CC   -!- FUNCTION: SODIUM-DEPENDENT GLUCOSE TRANSPORTER. HAS A NA+ TO
CC       GLUCOSE COUPLING RATIO OF 1:1.
CC   -!- FUNCTION: EFFICIENT SUBSTRATE TRANSPORT IN MAMMALIAN KIDNEY IS
CC       PROVIDED BY THE CONCERTED ACTION OF A LOW AFFINITY HIGH CAPACITY
CC       AND A HIGH AFFINITY LOW CAPACITY NA(+)/GLUCOSE COTRANSPORTER
CC       ARRANGED IN SERIES ALONG KIDNEY PROXIMAL TUBULES.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: KIDNEY, IN PROXIMAL TUBULE S1 SEGMENTS.
CC   -!- DEVELOPMENTAL STAGE: APPEARS ON EMBRYONIC DAY 17 AND GRADUALLY
CC       INCREASES UNTIL DAY 19. DECREASES BETWEEN DAY 19 AND BIRTH.
CC   -!- PTM: GLYCOSYLATED AT A SINGLE SITE.
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U29881; AAC52325.1; -.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   TIGRFAMs; TIGR00813; sss; 1.
DR   PROSITE; PS00456; NA_SOLUT_SYMP_1; 1.
DR   PROSITE; PS00457; NA_SOLUT_SYMP_2; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
KW   Transport; Sugar transport; Transmembrane; Sodium transport; Symport;
KW   Glycoprotein.
FT   DOMAIN        1     23       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     24     42       POTENTIAL.
FT   DOMAIN       43     59       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     60     80       POTENTIAL.
FT   DOMAIN       81    100       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    101    121       POTENTIAL.
```

```
FT   DOMAIN      122    166       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    167    186       POTENTIAL.
FT   DOMAIN      188    203       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    204    224       POTENTIAL.
FT   DOMAIN      225    268       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    269    289       POTENTIAL.
FT   DOMAIN      290    312       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    313    332       POTENTIAL.
FT   DOMAIN      333    421       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    422    441       POTENTIAL.
FT   DOMAIN      442    453       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    454    474       POTENTIAL.
FT   DOMAIN      475    524       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    525    545       POTENTIAL.
FT   DOMAIN      546    648       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    649    669       POTENTIAL.
FT   CARBOHYD    248    248       N-LINKED (GLCNAC. . .) (PROBABLE).
SQ   SEQUENCE   670 AA;  72961 MW;  0609562861618BB3 CRC64;

 Query Match             10.4%;  Score 308;  DB 1;  Length 670;
 Best Local Similarity   23.3%;  Pred. No. 5.5e-13;
 Matches  148;  Conservative   95;  Mismatches  209;  Indels  184;  Gaps   28;

Qy       8 LIAIIVFYLLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY 67
           :: |  ::||:: ||:|: :|| | |:          : || :     | :: |: :| |:
Db      24 ILVIAAYFLLVIGVGLWSMFRT-NRGTV----GGYFLAGRSMVWWPVGASLFASNIGSGH 78

Qy      68 INGTAEAVYVPGYGLAWAQAPIGYS-----LSLILGGLFFAKPMRSKGYVTMLDPFQQIY 122
             | |      | | |   |  |:      : |:|| ||    : : | :||     |
Db      79 FVGLA------GTGAASGLAVAGFEWNALFVVLLLGWLFVPVYL-TAGVITM----PQYL 127

Qy     123 GKRMGG-----------LLFIPALMGEMFWAAAIF--SALGATISVIIDVDMHISVIIS 168
            || ||           |:|   :    :: |:|   |||  |        : |||
Db     128 RKRFGGRRIRLYLSVLSLFLYIFTKISVDMFSGAVFIQQALGWNI--------YASVIAL 179

Qy     169 ALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY---- 224
             |  :||: ||| :: ||| || | |  | :|    :|         |:: :  ||
Db     180 LGITMIYTVTGGLAALMYTDTVQTFVILAGAFILTGYAFHEVG----GYSGLFDKYLGAV 235

Qy     225 ------QKPWLGTVDSSEVYSWLDSFLLL---MLGGIPW-----------------QAY 257
                : | :| : |:      ||: ||   : ||:||                    |
Db     236 TSLTVSKDPAVGNISSTCYQPRPDSYHLLRDPVTGGLPWPALLLGLTIVSGWHWCSDQVI 295

Qy     258 FQRVLSSSSATYAQ----VLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKTT 313
            || |:  : |: :    :  :|      :| :| ::          ::  |  ||
Db     296 VQRCLAGKNLTHIKAGCILCGYLKLMPMFLMVMPGMI----------SRILY--PD---- 339

Qy     314 EEADMILPIVLQYLC-----------------PVYISFFGLGAVSAAVMSSADSSILS 354
            |   ::| | : :|                 |  :    |  : ||:||| |   |
Db     340 -EVACVVPEVCKRVCGTEVGCSNIAYPRLVVKLMPNGLRGLMLAVMLAALMSSLASIFNS 398

Qy     355 ASSMFARNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYG---LWYLSSD 411
             :|::|  :||    |  | |:|:: | |: |    | : |   : :   |       |: |
Db     399 SSTLFTMDIY-TRLRPRAGDRELLLVGRLWVVFIVAVSVAWLPVVQAAQGGQLFDYIQSV 457

Qy     412 LVYIV--IFPQLLCVLFVKGTNTYGAVAGYVSGLFLRI----------TGG--EP----- 452
             |:   :    :  |||   |  ||  | : || : :          ||    |
Db     458 SSYLAPPVSAVFVLALFVPRVNEKGAFWGLIGGLLMGLARLIPEFFFGTGSCVRPSACPA 517

Qy     453 ------YLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGT 506
                 |||   ::|:                 | ||| ::     |   :| : :|
Db     518 IFCRVHYLYFAIILFFCS------------GFLTLA-ISRCTAPIPQKHLHRLVFS--- 560

Qy     507 LPPKLDVFDAVVARHSEENMDKTILVKNENIKLDEL 542
                      |||:|       : |::  :||
Db     561 -----------LRHSKE--------EREDLDAEEL 576


RESULT 3
SL51_HUMAN
ID   SL51_HUMAN     STANDARD;      PRT;   664 AA.
AC   P13866;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sodium/glucose cotransporter 1 (Na(+)/glucose cotransporter 1)
DE   (High affinity sodium-glucose cotransporter).
GN   SLC5A1 OR SGLT1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89345544; PubMed=2490366;
RA   Hediger M.A., Turk E., Wright E.M.;
RT   "Homology of the human intestinal Na+/glucose and Escherichia coli
RT   Na+/proline cotransporters.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:5748-5752(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Swan M.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20057165; PubMed=10591208;
RA   Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,
RA   Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
RA   Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA   Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
RA   Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA   Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
RA   Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA   Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
RA   Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA   Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
RA   Hall R.E., Hall-Tamlyn G., Heathcott R.W., Ho S., Holmes S.,
RA   Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
RA   Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA   Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
RA   Mcclay J., Mclaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA   Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA   Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
RA   Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA   Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
RA   Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA   Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA   Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA   Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA   Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
RA   Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA   Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA   Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
RA   Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA   Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA   Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
RA   Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA   Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA   Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
RA   Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,
RA   Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
RA   Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,
RA   McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,
RA   Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard M., Kedra D.,
RA   Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P.,
RA   Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
RA   Tilahun Y., Wright H.;
RT   "The DNA sequence of human chromosome 22.";
RL   Nature 402:489-495(1999).
RN   [4]
RP   VARIANT GGM ASN-28.
RX   MEDLINE=91179516; PubMed=2008213;
RA   Turk E., Zabel B., Mundlos S., Dyer J., Wright E.M.;
RT   "Glucose/galactose malabsorption caused by a defect in the
RT   Na+/glucose cotransporter.";
RL   Nature 350:354-356(1991).
```

```
RN   [5]
RP   VARIANT GGM GLY-28.
RX   PubMed=8195156;
RA   Turk E., Martin M.G., Wright E.M.;
RT   "Structure of the human Na+/glucose cotransporter gene SGLT1.";
RL   J. Biol. Chem. 269:15204-15209(1994).
CC   -!- FUNCTION: ACTIVELY TRANSPORTS GLUCOSE INTO CELLS BY NA+
CC       CO-TRANSPORT WITH A NA+ TO GLUCOSE COUPLING RATIO OF 2:1.
CC   -!- FUNCTION: EFFICIENT SUBSTRATE TRANSPORT IN MAMMALIAN KIDNEY IS
CC       PROVIDED BY THE CONCERTED ACTION OF A LOW AFFINITY HIGH CAPACITY
CC       AND A HIGH AFFINITY LOW CAPACITY NA(+)/GLUCOSE COTRANSPORTER
CC       ARRANGED IN SERIES ALONG KIDNEY PROXIMAL TUBULES.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Expressed mainly in intestine and kidney.
CC   -!- DISEASE: Defects in SLC5A1 are the cause of the intestinal
CC       monosaccharide transporter deficiency known as congenital glucose-
CC       galactose malabsorption (GGM). It is an autosomal recessive
CC       disorder manifesting itself within the first weeks of life. It is
CC       characterized by severe diarrhea and dehydration which are usually
CC       fatal unless glucose and galactose are eliminated from the diet.
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L29339; AAB59448.1; -.
DR   EMBL; L29328; AAB59448.1; JOINED.
DR   EMBL; L29330; AAB59448.1; JOINED.
DR   EMBL; L29329; AAB59448.1; JOINED.
DR   EMBL; L29331; AAB59448.1; JOINED.
DR   EMBL; L29332; AAB59448.1; JOINED.
DR   EMBL; L29333; AAB59448.1; JOINED.
DR   EMBL; L29334; AAB59448.1; JOINED.
DR   EMBL; L29335; AAB59448.1; JOINED.
DR   EMBL; L29336; AAB59448.1; JOINED.
DR   EMBL; L29337; AAB59448.1; JOINED.
DR   EMBL; L29338; AAB59448.1; JOINED.
DR   EMBL; M24847; AAA60320.1; -.
DR   EMBL; AL022321; CAA18445.2; -.
DR   EMBL; Z83849; CAB06090.2; -.
DR   EMBL; Z74021; CAA98336.2; -.
DR   EMBL; Z80998; CAB02632.2; -.
DR   EMBL; Z83839; CAB06087.1; -.
DR   EMBL; Z74021; CAB06087.1; JOINED.
DR   EMBL; Z80998; CAB06087.1; JOINED.
DR   EMBL; Z83849; CAB06087.1; JOINED.
DR   EMBL; AL022321; CAB06087.1; JOINED.
DR   PIR; A33545; A33545.
DR   Genew; HGNC:11036; SLC5A1.
DR   MIM; 182380; -.
DR   MIM; 606824; -.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   TIGRFAMs; TIGR00813; sss; 1.
DR   PROSITE; PS00456; NA_SOLUT_SYMP_1; 1.
DR   PROSITE; PS00457; NA_SOLUT_SYMP_2; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
KW   Transport; Sugar transport; Transmembrane; Sodium transport; Symport;
KW   Glycoprotein; Disease mutation.
FT   DOMAIN        1     28       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     29     47       POTENTIAL.
FT   DOMAIN       48     64       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     65     85       POTENTIAL.
FT   DOMAIN       86    105       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    106    126       POTENTIAL.
FT   DOMAIN      127    171       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    172    191       POTENTIAL.
FT   DOMAIN      193    208       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    209    229       POTENTIAL.
FT   DOMAIN      230    270       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    271    291       POTENTIAL.
FT   DOMAIN      292    314       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    315    334       POTENTIAL.
FT   DOMAIN      335    423       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    424    443       POTENTIAL.
FT   DOMAIN      444    455       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    456    476       POTENTIAL.
FT   DOMAIN      477    526       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    527    547       POTENTIAL.
FT   DOMAIN      548    642       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    643    663       POTENTIAL.
FT   CARBOHYD    248    248       N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT   VARIANT      28     28       D -> G (IN GGM).
FT                                /FTId=VAR_013630.
FT   VARIANT      28     28       D -> N (IN GGM).
FT                                /FTId=VAR_007168.
SQ   SEQUENCE   664 AA;  73497 MW;  2B403376595EAB74 CRC64;

  Query Match             10.3%;  Score 306;  DB 1;  Length 664;
  Best Local Similarity   22.8%;  Pred. No. 7.4e-13;
  Matches  148;  Conservative  104;  Mismatches  218;  Indels  178;  Gaps   30;

Qy      11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING 70
           |::::::::: ||:|| : | | |:         : || :     | :: |: :| |:  |
Db      32 IVIYFVVVMAVGLWAMFST-NRGTV----GGFFLAGRSMVWWPIGASLFASNIGSGHFVG 86

Qy      71 TAEAVYVPGYGLAWAQAPIGYS-----LSLILGGLFFAKPMRSK-GYVTMLDPFQQIYGK 124
           |     | | |   |  |:      | ::|| ||   |:  | | |||     :   |
Db      87 LA------GTGAASGIAIGGFEWNALVLVVVLGWLFV--PIYIKAGVVTM----PEYLRK 134

Qy     125 RMGG------------LLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIA 172
           | ||             ||:|   :    :: |||          |:: : ::::::: :   |
Db     135 RFGGQRIQVYLSLLSLLLYIFTKISADIFSGAIF------INLALGLNLYLAIFLLLAIT 188

Qy     173 TLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQK--PWL- 229
           |||: ||| :| ||| :|   : ||  |   ||              |: |   || |  | :
Db     189 ALYTITGGLAAVIYTDTLQTVIMLVGSLILTGFAFHEVG----GYDAFMEKYMKAIPTIV 244

Qy     230 ---GTVDSSEVYS-WLDSFLLL---MLGGIPW-----------------QAYFQRVLSS 264
               |   : |:   ||| :    : | :||                 |   || ||:
Db     245 SDGNTTFQEKCYTPRADSFHIFRDPLTGDLPWPGFIFGMSILTLWYWCTDQVIVQRCLSA 304

Qy     265 SSATYAQ----VLSFLAAFGCLVMAIPAIL--------IGAI---------GASTDWNQT 303
           : :: :   :  :|      :| :| ::          |  :         |
Db     305 KNMSHVKGGCILCGYLKLMPMFIMVMPGMISRILYTEKIACVVPSECEKYCGTKVGCTNI 364

Qy     304 AYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNI 363
           ||              | ::  | |   :    |  : |::||| |  |||::|  :|
Db     365 AY--------------PTLVVELMPNGLRGLMLSVMLASLMSSLTSIFNSASTLFTMDI 409

Qy     364 YQLSFRQNASDKEIVWVMRITVFV-FGASATAMALLTKTVYG--LWYLSSDLVYI--VIF 418
           |    |: ||:|||::   |: : |  | |   : ::     |    |: |   |:   |
Db     410 Y-AKVRKRASEKELMIAGRLFILVLIGISIAWVPIVQSAQSGQLFDYIQSITSYLGPPIA 468

Qy     419 PQLLCVLFVKGTNTYGAVAGYVSGLFLRI----------TG------------GEPYLY 455
             |  :| |  |  ||  | : || : |           ||              |  |||
Db     469 AVFLLAIFWKRVNEPGAFWGLILGLLIGISRMITEFAYGTGSCMEPSNCPTIICGVHYLY 528

Qy     456 LQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFD 515
           ::|                                 ||:| : || | | :          | : ::
Db     529 FAIILF----------------------AISFITIVVISLLTKPI-------PDVHLYR 558

Qy     516 AV--VARHSEENMDKTILVKNENIKLDELALVKPRQSMTLSSTFTNKE 561
           :    || :|  :  : |||:        |:::: : :    |:
Db     559 LCWSLRNSKEERID--LDAEEENIQ------EGPKETIEIETQVPEKK 598


RESULT 4
SL51_RAT
ID   SL51_RAT       STANDARD;      PRT;   665 AA.
AC   P53790; P97787;
```

```
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Sodium/glucose cotransporter 1 (Na(+)/glucose cotransporter 1)
DE   (High affinity sodium-glucose cotransporter).
GN   SLC5A1 OR SGLT1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Kidney;
RX   MEDLINE=94216314; PubMed=8163506;
RA   Lee W.S., Kanai Y., Wells R.G., Hediger M.A.;
RT   "The high affinity Na+/glucose cotransporter. Re-evaluation of
RT   function and distribution of expression.";
RL   J. Biol. Chem. 269:12032-12039(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kasahara M., Mori K.;
RL   Submitted (MAY-1993) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Jejunum;
RA   Aoshima H., Yokoyama T., Tanizaki J., Izu H., Yamada M.;
RT   "The sugar specificity of Na/glucose cotransporter from rat jejunum.";
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ACTIVELY TRANSPORTS GLUCOSE INTO CELLS BY NA+
CC       CO-TRANSPORT WITH A NA+ TO GLUCOSE COUPLING RATIO OF 2:1.
CC   -!- FUNCTION: EFFICIENT SUBSTRATE TRANSPORT IN MAMMALIAN KIDNEY IS
CC       PROVIDED BY THE CONCERTED ACTION OF A LOW AFFINITY HIGH CAPACITY
CC       AND A HIGH AFFINITY LOW CAPACITY NA(+)/GLUCOSE COTRANSPORTER
CC       ARRANGED IN SERIES ALONG KIDNEY PROXIMAL TUBULES.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- DEVELOPMENTAL STAGE: APPEARS ON EMBRYONIC DAY 18 AND GRADUALLY
CC       INCREASES UNTIL BIRTH.
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U03120; AAA19015.1; -.
DR   EMBL; D16101; BAA03676.1; -.
DR   EMBL; AB000729; BAA19172.1; -.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   TIGRFAMs; TIGR00813; sss; 1.
DR   PROSITE; PS00456; NA_SOLUT_SYMP_1; 1.
DR   PROSITE; PS00457; NA_SOLUT_SYMP_2; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
KW   Transport; Sugar transport; Transmembrane; Sodium transport; Symport;
KW   Glycoprotein.
FT   DOMAIN       1     28       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     29    47       POTENTIAL.
FT   DOMAIN       48    64       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     65    85       POTENTIAL.
FT   DOMAIN       86    105      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     106   126      POTENTIAL.
FT   DOMAIN       127   171      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     172   191      POTENTIAL.
FT   DOMAIN       193   208      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     209   229      POTENTIAL.
FT   DOMAIN       230   270      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     271   291      POTENTIAL.
FT   DOMAIN       292   314      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     315   334      POTENTIAL.
FT   DOMAIN       335   423      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     424   443      POTENTIAL.
FT   DOMAIN       444   455      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     456   476      POTENTIAL.
FT   DOMAIN       477   526      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     527   547      POTENTIAL.
FT   DOMAIN       548   643      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     644   664      POTENTIAL.
FT   CARBOHYD     248   248      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     354   354      Y -> H (IN REF. 3).
SQ   SEQUENCE   665 AA;  73066 MW;  A92038D964BFF061 CRC64;
```

```
  Query Match             10.3%;  Score 306;  DB 1;  Length 665;
  Best Local Similarity   23.5%;  Pred. No. 7.4e-13;
  Matches  155;  Conservative  105;  Mismatches  242;  Indels  158;  Gaps   29;

Qy      11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING 70
           |::::::::: ||:|| : | | |:         : || :     | :: |: :| |:  |
Db      32 IVIYFVVVMAVGLWAMFST-NRGTV----GGFFLAGRSMVWWPIGASLFASNIGSGHFVG 86

Qy      71 TAEAVYVPGYGLAWAQAPIGYSLS-----LILGGLFFAKPMRSK-GYVTMLDPFQQIYGK 124
           |      | | |   |  |:  :     ::|| ||   |:  | | |||     :   |
Db      87 LA------GTGAAAGIAMGGFEWNALVFVVVLGWLFV--PIYIKAGVVTM----PEYLRK 134

Qy     125 RMGG-----------LLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIA 172
           | ||           ||:|  :   :: |||      |:: : :|:::: |   |
Db     135 RFGGKRIQIYLSVLSLLLYIFTKISADIFSGAIF------INLALGLDIYLAIFILLAIT 188

Qy     173 TLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQK--PWL- 229
           |||: ||| :| ||| :|   : || :|   ||             |: |   || |  | |
Db     189 ALYTITGGLAAVIYTDTLQTAIMLVGSFILTGFAFREVG----GYEAFMDKYMKAIPTLV 244

Qy     230 --GTVD-SSEVYS-WLDSFLLL---MLGGIPW-----------------QAYFQRVLSS 264
             | :    | |:   ||| :    : | :||                 |   || ||:
Db     245 SDGNITVKEECYTPRADSFHIFRDPITGDMPWPGLIFGLSILALWYWCTDQVIVQRCLSA 304

Qy     265 SSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDP------KTTEEADM 318
           : :: :    |  :  |:     :: | |      :: |  ||              ::
Db     305 KNMSHVKAGCTLCGYLKLLPMFLMVMPGMISRILYTDKIACVLPSECKKYCGTPVGCTNI 364

Qy     319 ILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIV 378
           | ::  | |  :    |  : |::|||  |   |||::|  :||    |: ||:||::
Db     365 AYPTLVVELMPNGLRGLMLSVMMASLMSSLTSIFNSASTLFTMDIY-TKIRKGASEKELM 423

Qy     379 WVMRITVFV-FGASATAMALLTKTVYG--LWYLSSDLVYI--VIFPQLLCVLFVKGTNTY 433
             |: : |  | |  : ::     |   |: |  |:  |    |  :| | |
Db     424 IAGRLFILVLIGISIAWVPIVQSAQSGQLFDYIQSITSYLGPPIAAVFLLAIFCKRVNEP 483

Qy     434 GAVAGYVSGLFLRI----------TG-------------GEPYLYLQPLIFYPGYYPDDN 470
           ||  | : |  : |           ||             |  |||   ::|
Db     484 GAFWGLILGFLIGISRMITEFAYGTGSCMEPSNCPKIICGVHYLYFAIILF--------- 534

Qy     471 GIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAV--VARHSEENMDK 528
                          | :| : || | | :          | : ::     :   :|| :|
Db     535 -------------AISVVTVLVISLLTKPI-------PDVHLYRLCWSLRNSTEERID- 572

Qy     529 TILVKNENIKLDELALVKPRQSMTLSSTFTNKE-----------AFLDVDSSPEGSGTED 577
           |   |   ::|     |: :: : :    ||                || |  |: :  |:
Db     573 --LDAGEEEPVEE----DPKDTIEIDAEAPQKEKGCFRKAYDLFCGLDQDKGPKMTKEEE 626
```

```
RESULT 5
SL54_PIG
ID   SL54_PIG       STANDARD;      PRT;   660 AA.
AC   P31636;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Low affinity sodium-glucose cotransporter (Sodium/glucose
DE   cotransporter 3) (Na(+)/glucose cotransporter 3).
GN   SLC5A4 OR SGLT3 OR SAAT1.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
```

```
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=93131881; PubMed=8420925;
RA   Kong C.-T., Yet S.-F., Lever J.E.;
RT   "Cloning and expression of a mammalian Na+/amino acid cotransporter
RT   with sequence similarity to Na+/glucose cotransporters.";
RL   J. Biol. Chem. 268:1509-1512(1993).
RN   [2]
RP   FUNCTION.
RX   MEDLINE=94357885; PubMed=8077195;
RA   McKenzie B., Panayotova-Heiermann M., Loo D.D.F., Lever J.E.,
RA   Wright E.M.;
RT   "SAAT1 is a low affinity Na+/glucose cotransporter and not an amino
RT   acid transporter. A reinterpretation.";
RL   J. Biol. Chem. 269:22488-22491(1994).
CC   -!- FUNCTION: SODIUM-DEPENDENT GLUCOSE TRANSPORTER. HAS A NA+ TO
CC       GLUCOSE COUPLING RATIO OF 1:1.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: KIDNEY, INTESTINE, LIVER, SKELETAL MUSCLE,
CC       AND SPLEEN.
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
CC   -!- CAUTION: WAS ORIGINALLY (REF.1) THOUGHT TO BE A SODIUM/NEUTRAL
CC       AMINO ACID COTRANSPORTER (SYSTEM A NEUTRAL AMINO ACID TRANSPORTER)
CC       RESPONSIBLE FOR THE SODIUM-DEPENDENT INTAKE OF NEUTRAL AMINO ACIDS
CC       SUCH AS ALANINE, GLYCINE, SERINE, CYSTEINE, AND PROLINE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L02900; AAC37325.1; -.
DR   PIR; A44432; A44432.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   TIGRFAMs; TIGR00813; sss; 1.
DR   PROSITE; PS00456; NA_SOLUT_SYMP_1; 1.
DR   PROSITE; PS00457; NA_SOLUT_SYMP_2; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
KW   Transport; Sugar transport; Transmembrane; Sodium transport; Symport;
KW   Glycoprotein.
FT   DOMAIN        1     28       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      29    47       POTENTIAL.
FT   DOMAIN        48    64       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      65    85       POTENTIAL.
FT   DOMAIN        86    105      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      106   126      POTENTIAL.
FT   DOMAIN        127   171      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      172   191      POTENTIAL.
FT   DOMAIN        192   208      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      209   229      POTENTIAL.
FT   DOMAIN        230   270      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      271   291      POTENTIAL.
FT   DOMAIN        292   314      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      315   334      POTENTIAL.
FT   DOMAIN        335   423      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      424   443      POTENTIAL.
FT   DOMAIN        444   455      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      456   476      POTENTIAL.
FT   DOMAIN        477   526      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      527   547      POTENTIAL.
FT   DOMAIN        548   638      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      639   659      POTENTIAL.
FT   CARBOHYD      248   248      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   660 AA;  72745 MW;  38616367F8F18F1A CRC64;

 Query Match             10.2%;  Score 303.5;  DB 1;  Length 660;
 Best Local Similarity   23.2%;  Pred. No. 1.1e-12;

 Matches  141;  Conservative  103;  Mismatches  230;  Indels  135;  Gaps   26;

Qy      11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING 70
           |:::::::: ||:||  || | |:          : |||:     | :: |: :| |:  |
Db      32 IVIYFVVVMAVGLWAMLRT-NRGTV----GGFFLAGRDVTWWPMGASLFASNIGSGHFVG 86

Qy      71 TAEAVYVPGYGLA---WAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQIY-GKRM 126
             |      |  :|   |          | |:: | ||       | :|| :  :: : |||:
Db      87 LAGTGAASGIAIAAFEW-----NALLLLLVLGWFFVPIYIKAGVMTMPEYLRKRFGGKRL 141

Qy     127 GGLLFIPAL-------MGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVG 179
             | | :|         :    :: |||      | : : :|:::::     |  :||: |
Db     142 QIYLSILSLFICVALRISSDIFSGAIF------IKLALGLDLYLAIFSLLAITAIYTITG 195

Qy     180 GLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQK--PWLGTVD---- 233
           || || ||| :|   : :| :| : ||     |   |: :  ||    | :  |
Db     196 GLASVIYTDTLQTIIMLIGSFILMGFAF----VEVGGYESFTEKYMNAIPTIVEGDNLTI 251

Qy     234 SSEVYS-WLDSFLLL---MLGGIPW-----------------QAYFQRVLSSSSATYAQ 271
           | : |:   ||| :    : | |||                 |   || ||    :: :
Db     252 SPKCYTPQGDSFHIFRDAVTGDIPWPGMIFGMTVVAAWYWCTDQVIVQRCLSGKDMSHVK 311

Qy     272 VLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKT----TEE--ADMILPIVLQ 325
             :  : |:     :: | |       : |  :|       ||   ::   |:::
Db     312 AACIMCGYLKLLPMFLMVMPGMISRILYTEKVACVVPSECVKHCGTEVGCSNYAYPLLVM 371

Qy     326 YLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITV 385
           | |  :    |  : |::|||  |   |||::|  ::|   |: ||:||::   |: :
Db     372 ELMPSGLRGLMLSVMLASLMSSLTSIFNSASTLFTMDLY-TKIRKQASEKELLIAGRLFI 430

Qy     386 FVFGASATAMALLTKTVYG---LWYLSSDLVYI--VIFPQLLCVLFVKGTNTYGA----V 436
             :   :    | :       |: |   |:   |    |  :| |  |  ||    :
Db     431 ILLIVISIVWVPLVQVAQNGQLFHYIESISSYLGPPIAAVFLLAIFCKRVNEQGAFWGLI 490

Qy     437 AGYVSGL------FLRITG-----------GEPYLYLQPLIFYPGYYPDDNGIYNQKF 477
           |:| ||      |:  ||             |  |||   ::|:
Db     491 IGFVMGLIRMIAEFVYGTGSCLAASNCPQIICGVHYLYFALILFF--------------- 535

Qy     478 PFKTLAMVTSFLTNICISYLAK-----YLFE-----SGTLPPKLDVFDAVVARH------ 521
                 | |  : || | |     :|:        :   ::|: ||   ||
Db     536 --------VSILVVLAISLLTKPIPDVHLYRLCWALRNSTEERIDL-DAEEKRHEEAHDG 586

Qy     522 -SEENMDKT 529
            |:| ::|
Db     587 VDEDNPEET 595


RESULT 6
SL52_RABIT
ID   SL52_RABIT     STANDARD;      PRT;   672 AA.
AC   P26430;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Sodium/nucleoside cotransporter (Na(+)/nucleoside cotransporter).
GN   SLC5A2 OR SNST1.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=92156077; PubMed=1740408;
RA   Pajor A.M., Wright E.M.;
RT   "Cloning and functional expression of a mammalian Na+/nucleoside
RT   cotransporter. A member of the SGLT family.";
RL   J. Biol. Chem. 267:3557-3560(1992).
CC   -!- FUNCTION: ACTIVELY TRANSPORTS URIDINE INTO CELLS BY NA+
CC       CO-TRANSPORT. MAY PLAY A ROLE IN REABSORPTION OF NUCLEOSIDES FROM
CC       GLOMERULAR FILTRATE BY THE PROXIMAL TUBULE IN KIDNEY, AND IN THE
CC       REGULATION OF CARDIAC CONTRACTILITY BY ADENOSINE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
```

```
CC   -!- TISSUE SPECIFICITY: MORE ABUNDANT IN HEART THAN IN KIDNEY, WHERE
CC       IT IS ABSENT FROM THE OUTER CORTEX.
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M84020; AAA31421.1; -.
DR   PIR; A42251; A42251.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   TIGRFAMs; TIGR00813; sss; 1.
DR   PROSITE; PS00456; NA_SOLUT_SYMP_1; 1.
DR   PROSITE; PS00457; NA_SOLUT_SYMP_2; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
KW   Transport; Transmembrane; Sodium transport; Symport; Glycoprotein.
FT   DOMAIN        1     25       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     26     44       POTENTIAL.
FT   DOMAIN       45     61       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     62     82       POTENTIAL.
FT   DOMAIN       83    102       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    103    123       POTENTIAL.
FT   DOMAIN      124    168       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    169    188       POTENTIAL.
FT   DOMAIN      189    205       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    206    226       POTENTIAL.
FT   DOMAIN      227    270       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    271    291       POTENTIAL.
FT   DOMAIN      292    314       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    315    334       POTENTIAL.
FT   DOMAIN      335    422       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    423    442       POTENTIAL.
FT   DOMAIN      443    454       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    455    475       POTENTIAL.
FT   DOMAIN      476    525       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    526    546       POTENTIAL.
FT   DOMAIN      547    650       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    651    671       POTENTIAL.
FT   CARBOHYD    250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   672 AA;  73161 MW;  E2D987B03B9C57B4 CRC64;

  Query Match             10.0%;  Score 298;  DB 1;  Length 672;
  Best Local Similarity   25.0%;  Pred. No. 2.5e-12;
  Matches  153;  Conservative   89;  Mismatches  232;  Indels  138;  Gaps   25;

Qy       9 IAII-VFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY 67
           ||:|  ::||:: ||:|:  || | |:        : || :      | :: |: :| |:
Db      26 IAVIAAYFLLVIGVGLWSMCRT-NRGTV----GGYFLAGRSMVWWPVGASLFASNIGSGH 80

Qy      68 INGTAEAVYVPGYGLAWAQAPIGYSLS----LILGGLFFAKPMRSKGYVTMLDPFQQIYG 123
             | |      | | |   |  |:  :    ::| |  ||    : | :||     |
Db      81 FVGLA------GTGAANGLAVAGFEWNALFVVLLLGWLFAPVYLTAGVITM----PQYLR 130

Qy     124 KRMGG-----------LLFIPALMGEMFWAAAIF--SALGATISVIIDVDMHISVIISA 169
           || ||             |:|   :    :: |:|   |||  |        : |||
Db     131 KRFGGHRIRLYLSVLSLFLYIFTKISVDMFSGAVFIQQALGWNI--------YASVIALL 182

Qy     170 LIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY----- 224
             |  :||: ||| :: ||| || | |  | :|   :|         |:: :  ||
Db     183 GITMVYTVTGGLAALMYTDTVQTFVIIAGAFILTGYAFHEVG----GYSGLFDKYMGAMT 238

Qy     225 -----QKPWLGTVDSSEVYSWLDSFLLL---MLGGIPW-----------------QAYF 258
                : | :| : ||      ||: ||   : | :||                    |
Db     239 SLTVSEDPAVGNISSSCYRPRPDSYHLLRDPVTGDLPWPALLLGLTIVSGWYWCSDQVIV 298

Qy     259 QRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKT----TE 314
           || |:  : |: :    |  :  |      :: | |      :: |   |:       ||
Db     299 QRCLAGRNLTHIKAGCILCGYLKLTPMFLMVMPGMISRILYPDEVACVAPEVCKRVCGTE 358

Qy     315 E--ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNA 372
           :::  | ::  | |  :    |  : ||:|||  |   |:|::|  :|| |  |  |
Db     359 VGCSNIAYPRLVVKLMPNGLRGLMLAVMLAALMSSLASIFNSSSTLFTMDIYTL--RPRA 416

Qy     373 SDKEIVWVMRITVFVFGASATAMALLTKTVYG---LWYLSSDLVYIV--IFPQLLCVLFV 427
             : |:: | |: |    | : |   : :   |      |: |   |:   :    : |||
Db     417 GEGELLLVGRLWVVFIVAVSVAWLPVVQAAQGGQLFDYIQSVSSYLAPPVSAVFVVALFV 476

Qy     428 KGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMV-- 485
             |  ||  | : || : :   |                                :| | | : |
Db     477 PRVNEKGAFWGLIGGLLMGLARLIP----------------------EFSFGTGSCVRP 513

Qy     486 ---TSFLTNICISYLAKYLFE-SG---------TLP-PKLDVFDAVVA-RHSEENMDKTI 530
               :||  :   | |  ||  ||          | | |:  :   | : |||:|
Db     514 SACPAFLCRVHYLYFAIVLFFCSGLLIIIVSLCTAPIPRKHLHRLVFSLRHSKE------ 567

Qy     531 LVKNENIKLDEL 542
             : |::  |||
Db     568 --EREDLDADEL 577


RESULT 7
SL54_HUMAN
ID   SL54_HUMAN     STANDARD;      PRT;   659 AA.
AC   Q9NY91; O15279;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Low affinity sodium-glucose cotransporter (Sodium/glucose
DE   cotransporter 3) (Na(+)/glucose cotransporter 3).
GN   SLC5A4 OR SAAT1 OR SGLT2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Small intestine;
RA   Gorboulev V., Baumgarten K., Veyhl M., Koepsell H.;
RT   "The molecular cloning and functional characterization of the human
RT   SGLT2 transporter.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20057165; PubMed=10591208;
RA   Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,
RA   Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
RA   Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA   Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
RA   Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA   Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
RA   Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA   Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
RA   Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA   Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
RA   Hall R.E., Hall-Tamlyn G., Heathcott R.W., Ho S., Holmes S.,
RA   Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
RA   Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA   Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
RA   Mcclay J., Mclaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA   Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA   Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
RA   Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA   Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
RA   Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA   Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA   Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA   Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA   Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
RA   Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA   Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA   Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
```

```
RA   Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA   Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA   Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
RA   Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA   Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA   Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
RA   Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,
RA   Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
RA   Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,
RA   McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,
RA   Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard M., Kedra D.,
RA   Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P.,
RA   Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
RA   Tilahun Y., Wright H.;
RT   "The DNA sequence of human chromosome 22.";
RL   Nature 402:489-495(1999).
RN   [3]
RP   SEQUENCE OF 73-247 FROM N.A.
RC   TISSUE=Brain;
RA   Poppe R., Koepsell H.;
RL   Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: SODIUM-DEPENDENT GLUCOSE TRANSPORTER (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ133127; CAB81772.1; -.
DR   EMBL; AL008723; CAB51758.1; -.
DR   EMBL; U41897; AAB61732.1; -.
DR   Genew; HGNC:11039; SLC5A4.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   TIGRFAMs; TIGR00813; sss; 1.
DR   PROSITE; PS00456; NA_SOLUT_SYMP_1; 1.
DR   PROSITE; PS00457; NA_SOLUT_SYMP_2; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
KW   Transport; Sugar transport; Transmembrane; Sodium transport; Symport;
KW   Glycoprotein.
FT   DOMAIN        1     28       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      29    47       POTENTIAL.
FT   DOMAIN        48    64       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      65    85       POTENTIAL.
FT   DOMAIN        86    105      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      106   126      POTENTIAL.
FT   DOMAIN        127   171      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      172   191      POTENTIAL.
FT   DOMAIN        192   208      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      209   229      POTENTIAL.
FT   DOMAIN        230   270      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      271   291      POTENTIAL.
FT   DOMAIN        292   314      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      315   334      POTENTIAL.
FT   DOMAIN        335   423      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      424   443      POTENTIAL.
FT   DOMAIN        444   455      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      456   476      POTENTIAL.
FT   DOMAIN        477   526      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      527   547      POTENTIAL.
FT   DOMAIN        548   637      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      638   658      POTENTIAL.
FT   CARBOHYD      248   248      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT      76    76       A -> V (IN REF. 3).
FT   CONFLICT      106   106      S -> P (IN REF. 3).
FT   CONFLICT      243   243      V -> I (IN REF. 3).
SQ   SEQUENCE   659 AA;  72455 MW;  F8A34AED648B523A CRC64;
```

```
  Query Match             9.9%;  Score 294;  DB 1;  Length 659;
  Best Local Similarity   22.5%;  Pred. No. 4.4e-12;
  Matches  135;  Conservative   97;  Mismatches  225;  Indels  144;  Gaps   22;

Qy      11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING 70
           |:::::::: ||:||  :| | |:          : |||:     | :: |: :|  :  |
Db      32 IVIYFLVVMAVGLWAMLKT-NRGTI----GGFFLAGRDMAWWPMGASLFASNIGSNHYVG 86

Qy      71 TAEAVYVPGYGLAWAQAPIGYS-----LSLILGGLFFAKPMRSKGYVTMLDPFQQIYGKR 125
           |      | | |   | : :      : |||| :|    ::| | :||     :  ||
Db      87 LA------GTGAASGVATVTFEWTSSVMLLILGWIFVPIYIKS-GVMTM----PEYLKKR 135

Qy     126 MGG-----------LLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIAT 173
           ||              : :  |:    :| |||      | : : :|:::: |   :
Db     136 FGGERLQVYLSILSLFICVVLLISADIFAGAIF------IKLALGLDLYLAIFILLAMTA 189

Qy     174 LYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVD 233
           :||  ||| || ||| :|   : :| :| : || :       |: :   ||      |:
Db     190 VYTTTGGLASVIYTDTLQTIIMLIGSFILMGFAFNEVG----GYESFTEKYVNATPSVVE 245

Qy     234 ------SSEVYS-WLDSFLLL---MLGGIPW-----------------QAYFQRVLSSS 265
                |:  |:   ||| :    : | |||                 |   || |
Db     246 GDNLTISASCYTPRADSFHIFRDAVTGDIPWPGIIFGMPITALWYWCTNQVIVQRCLCGK 305

Qy     266 SATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLP---------DPKTTEEA 316
           :: :   : |:  |:     :: | |      :  |  :|          |   |  |
Db     306 DMSHVKAACIMCAYLKLLPMFLMVMPGMISRILYTDMVACVVPSECVKHCGVDVGCTNYA 365

Qy     317 DMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKE 376
           | ::  | |  :    |  : |::||| |   |||::|  ::|    |: ||:||
Db     366 ---YPTMVLELMPQGLRGLMLSVMLASLMSSLTSIFNSASTLFTIDLY-TKMRKQASEKE 421

Qy     377 IVWVMRITVFVFGASATAMALLTKTVYG---LWYLSSDLVYI--VIFPQLLCVLFVKGTN 431
           ::   || | :    :     | :       : |  |   |:   |    :  :| |  |
Db     422 LLIAGRIFVLLLTVVSIVWVPLVQVSQNGQLIHYTESISSYLGPPIAAVFVLAIFCKRVN 481

Qy     432 TYGAVAGYVSGLFL----------------------RITGGEPYLYLQPLIFYPGYYPD 468
             ||  | : || :                     :|  |  |||   ::|:
Db     482 EQGAFWGLMVGLAMGLIRMITEFAYGTGSCLAPSNCPKIICGVHYLYFSIVLFF------ 535

Qy     469 DNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDA--VVARHSEENM 526
                              | |  : || | | :       | : ::    |:   :|| :
Db     536 -----------------GSMLVTLGISLLTKPI-------PDVHLYRLCWVLRNSTEERI 571

Qy     527 D 527
           |
Db     572 D 572
```

```
RESULT 8
SL51_SHEEP
ID   SL51_SHEEP     STANDARD;      PRT;   664 AA.
AC   P53791;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Sodium/glucose cotransporter 1 (Na(+)/glucose cotransporter 1)
DE   (High affinity sodium-glucose cotransporter).
GN   SLC5A1 OR SGLT1.
OS   Ovis aries (Sheep).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Caprinae; Ovis.
OX   NCBI_TaxID=9940;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Jejunal mucosa;
RX   MEDLINE=96077158; PubMed=7492327;
RA   Tarpey P., Wood I.S., Shirazi-Beechey S.P., Beechey R.B.;
RT   "Amino acid sequence and the cellular location of the Na(+)-dependent
RT   D-glucose symporters (SGLT1) in the ovine enterocyte and the parotid
RT   acinar cell.";
RL   Biochem. J. 312:293-300(1995).
```

```
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=98050042; PubMed=9388688;
RA   Shillingford J.M., Wood I.S., Shennan D.B., Shirazi-Beechey S.P.,
RA   Beechey R.B.;
RT   "Determination of the sequence of a mRNA from lactating sheep mammary
RT   gland that encodes a protein identical to the Na(+)-dependent glucose
RT   transporter (SGLT1).";
RL   Biochem. Soc. Trans. 25:467-467(1997).
CC   -!- FUNCTION: ACTIVELY TRANSPORTS GLUCOSE INTO CELLS BY NA+
CC       CO-TRANSPORT WITH A NA+ TO GLUCOSE COUPLING RATIO OF 2:1.
CC   -!- FUNCTION: EFFICIENT SUBSTRATE TRANSPORT IN MAMMALIAN KIDNEY IS
CC       PROVIDED BY THE CONCERTED ACTION OF A LOW AFFINITY HIGH CAPACITY
CC       AND A HIGH AFFINITY LOW CAPACITY NA(+)/GLUCOSE COTRANSPORTER
CC       ARRANGED IN SERIES ALONG KIDNEY PROXIMAL TUBULES.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X82411; CAA57809.1; -.
DR   EMBL; X82410; CAA57808.1; -.
DR   EMBL; AJ001026; CAA04483.1; -.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   TIGRFAMs; TIGR00813; sss; 1.
DR   PROSITE; PS00456; NA_SOLUT_SYMP_1; 1.
DR   PROSITE; PS00457; NA_SOLUT_SYMP_2; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
KW   Transport; Sugar transport; Transmembrane; Sodium transport; Symport;
KW   Glycoprotein.
FT   DOMAIN        1     28       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     29     47       POTENTIAL.
FT   DOMAIN       48     64       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     65     85       POTENTIAL.
FT   DOMAIN       86    105       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    106    126       POTENTIAL.
FT   DOMAIN      127    171       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    172    191       POTENTIAL.
FT   DOMAIN      193    208       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    209    229       POTENTIAL.
FT   DOMAIN      230    270       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    271    291       POTENTIAL.
FT   DOMAIN      292    314       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    315    334       POTENTIAL.
FT   DOMAIN      335    423       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    424    443       POTENTIAL.
FT   DOMAIN      444    455       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    456    476       POTENTIAL.
FT   DOMAIN      477    526       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    527    547       POTENTIAL.
FT   DOMAIN      548    642       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    643    663       POTENTIAL.
FT   CARBOHYD    248    248       N-LINKED (GLCNAC. .) (POTENTIAL).
SQ   SEQUENCE   664 AA;  73178 MW;  820AC019B5C93987 CRC64;

  Query Match             9.9%;  Score 294;  DB 1;  Length 664;
  Best Local Similarity   23.9%;  Pred. No. 4.5e-12;
  Matches  127;  Conservative   93;  Mismatches  202;  Indels  110;  Gaps   23;

Qy      11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY--- 67
           |:::::::: ||:|| : | | |:          : || :     | :: |: :| |:
Db      32 IVIYFVVVMAVGLWAMFST-NRGTV----GGFFLAGRSMVWWPIGASLFASNIGSGHFVG 86

Qy      68 INGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSK-GYVTMLDPFQQIYGKRM 126
           : || | :  | | |       | ::|| :|   |:  | | |||    :  ||

Db      87 LAGTGAAAGIATGGFEWN----ALILVVLLGWVFV--PIYIKAGVVTM----PEYLRKRF 136

Qy     127 GG------------LLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATL 174
           ||            :|:|   :    :: |||      |:: : :|:::::: |   | |
Db     137 GGQRIQVYLSVLSLVLYIFTKISADIFSGAIF------INLALGLDLYLAIFILLAITAL 190

Qy     175 YTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDS 234
           ||: ||| :| ||| :|   : :| :|   ||          |::|   ||    : || |
Db     191 YTITGGLAAVIYTDTLQTVIMLLGSFILTGFAFHEVG----GYSAFVTKYMNA-IPTVTS 245

Qy     235 -------SEVYS-WLDSFLLL---MLGGIPW-----------------QAYFQRVLSSS 265
                 | |:   ||| :    : | :||                   |   || ||:
Db     246 YGNTTVKKECYTPRADSFHIFRDPLKGDLPWPGLIFGLTIISLWYWCTDQVIVQRCLSAK 305

Qy     266 SATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKTTEE--------AD 317
           : :: :    :  :  |:     :: | |       : |  :  |   |:        :
Db     306 NMSHVKAGCIMCGYMKLLPMFLMVMPGMISRILFTEKVACTV--PSECEKYCGTKVGCTN 363

Qy     318 MILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEI 377
           :  | ::  | |  :    |  : |::|||  |   |||::|  :||    |: ||:||:
Db     364 IAYPTLVVELMPNGLRGLMLSVMLASLMSSLTSIFNSASTLFTMDIY-TKIRKKASEKEL 422

Qy     378 VWVMRITVFV-FGASATAMALLTKTVYG--LWYLSSDLVYI--VIFPQLLCVLFVKGTNT 432
           :   |: : |  | |   : ::     |    |: |   |:   |    |  :| |  |
Db     423 MIAGRLFMLVLIGVSIAWVPIVQSAQSGQLFDYIQSITSYLGPPIAAVFLLAIFCKRVNE 482

Qy     433 YGAVAGYVSGLFLRI----------TG------------GEPYLYLQPLIF 461
            ||  | : |  : :           ||            |  |||   ::|
Db     483 PGAFWGLIIGFLIGVSRMITEFAYGTGSCMEPSNCPTIICGVHYLYFAIILF 534


RESULT 9
SGLT_VIBPA
ID   SGLT_VIBPA     STANDARD;      PRT;   543 AA.
AC   P96169;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sodium/glucose cotransporter (Na(+)/glucose symporter).
GN   SGLT.
OS   Vibrio parahaemolyticus.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=670;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AQ3334;
RX   MEDLINE=96248401; PubMed=8652595;
RA   Sarker R.I., Okabe Y., Tsuda M., Tsuchiya T.;
RT   "Sequence of a Na+/glucose symporter gene and its flanking regions of
RT   Vibrio parahaemolyticus.";
RL   Biochim. Biophys. Acta 1281:1-4(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20400508; PubMed=10835424;
RA   Turk E., Kim O., Le Coutre J., Whitelegge J.P., Eskandari S.,
RA   Lam J.T., Kreman M., Zampighi G., Faull K.F., Wright E.M.;
RT   "Molecular characterization of Vibrio parahaemolyticus vSGLT: a model
RT   for sodium-coupled sugar cotransporters.";
RL   J. Biol. Chem. 275:25711-25716(2000).
RN   [3]
RP   MASS SPECTROMETRY OF FORMYLATED FORM, AND REVISIONS TO N-TERMINUS.
RX   MEDLINE=20222957; PubMed=10757971;
RA   le Coutre J., Whitelegge J.P., Gross A., Turk E., Wright E.M.,
RA   Kaback H.R., Faull K.F.;
RT   "Proteomics on full-length membrane proteins using mass
RT   spectrometry.";
RL   Biochemistry 39:4237-4242(2000).
CC   -!- FUNCTION: ACTIVELY TRANSPORTS GLUCOSE INTO CELLS BY NA+ CO-
CC       TRANSPORT (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- MASS SPECTROMETRY: MW=60680; METHOD=Electrospray.
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
CC   ---------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------------
DR   EMBL; D78137; BAA11215.1; ALT_FRAME.
DR   EMBL; AF255301; AAF80602.1; -.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   TIGRFAMs; TIGR00813; sss; 1.
DR   PROSITE; PS00456; NA_SOLUT_SYMP_1; 1.
DR   PROSITE; PS00457; NA_SOLUT_SYMP_2; FALSE_NEG.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
KW   Transport; Sugar transport; Transmembrane; Sodium transport; Symport.
FT   TRANSMEM     10     30       POTENTIAL.
FT   TRANSMEM     45     65       POTENTIAL.
FT   TRANSMEM     79     99       POTENTIAL.
FT   TRANSMEM    129    149       POTENTIAL.
FT   TRANSMEM    156    176       POTENTIAL.
FT   TRANSMEM    193    213       POTENTIAL.
FT   TRANSMEM    246    266       POTENTIAL.
FT   TRANSMEM    287    307       POTENTIAL.
FT   TRANSMEM    345    365       POTENTIAL.
FT   TRANSMEM    401    421       POTENTIAL.
FT   TRANSMEM    427    447       POTENTIAL.
FT   TRANSMEM    455    475       POTENTIAL.
FT   TRANSMEM    483    503       POTENTIAL.
FT   TRANSMEM    523    543       POTENTIAL.
SQ   SEQUENCE   543 AA;  58874 MW;  61BE3F7E380BC32C CRC64;

  Query Match             9.9%;  Score 293.5;  DB 1;  Length 543;
  Best Local Similarity   25.1%;  Pred. No. 3.9e-12;
  Matches  139;  Conservative   96;  Mismatches  198;  Indels  121;  Gaps   27;

Qy       4 HVEGLIAIIVF--YLLILL-VGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
Db       6 HGLSFIDIMVFAIYVAIIIGVGLWV---SRDKKGTQKSTEDYFLAGKSLPWWAVGASLIA 62

Qy      61 TWV---------GGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGY 111
Db      63 ANISAEQFIGMSGSGYSIGLAIASY------EWMSA-----ITLIIVGKYFLPIFIEKGI 111

Qy     112 VTMLDPFQQIYGKRMGGLLFIPALMGEMFWAAA-IFSAL------GATISVIIDVDMHI 163
Db     112 YTIPEFVEKRFNKKLKTILAV-------FWISLYIFVNLTSVLYLGGLALETILGIPLMY 164

Qy     164 SVIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAK 223
Db     165 SILGLALFALVYSIYGGLSAVVWTDVIQVFFLVLG-----GFMTTYMAVSFIGGT----- 214

Qy     224 YQKPWLGTV---------------DSSEVYSWLDSFLLLMLGGIPW----------QAY 257
Db     215 --DGWFAGVSKMVDAAPGHFEMILDQSNPQYMNLPG-IAVLIGGL-WVANLYYWGFNQYI 270

Qy     258 FQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIG-AIGASTDWNQTAYGLPDPKTTE-- 314
Db     271 IQRTLAAKSVSEAQKGIVFAAFLKLIVPFLVVLPGIAAYVITSDPQLMASLGDIAATNLP 330

Qy     315 ---EADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQN 371
Db     331 SAANADKAYPWLTQFL-PVGVKGVVFAALAAAIVSSLASMLNSTATIFTMDIYKEYISPD 389

Qy     372 ASDKEIVWVMRITVFVFGASATAMALLTKTVYG--LWYLSSDLVYI-----VIFPQLLCV 424
Db     390 SGDHKLVNV--------GRTAAVVALIIACLIAPMLGGIGQAFQYIQEYTGLVSPGILAV 441

Qy     425 ----LFVKGTNTYGAVAGYVSGLFLRITGGEPY-LYLQPLIFYPGYYP-DDNGIYNQKFP 478
Db     442 FLLGLFWKKTTSKGAIIGVVASI--------PFALFLK---FMPLSMPFMDQMLYTLLFT 490

Qy     479 FKTLAMVTSFLTNI 492
Db     491 MVVIAF-TSLSTSI 503


RESULT 10
SL52_HUMAN
ID   SL52_HUMAN     STANDARD;      PRT;   672 AA.
AC   P31639;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Sodium/glucose cotransporter 2 (Na(+)/glucose cotransporter 2)
DE   (Low affinity sodium-glucose cotransporter).
GN   SLC5A2 OR SGLT2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=93035768; PubMed=1415574;
RA   Wells R.G., Pajor A.M., Kanai Y., Turk E., Wright E.M., Hediger M.A.;
RT   "Cloning of a human kidney cDNA with similarity to the sodium-glucose
RT   cotransporter.";
RL   Am. J. Physiol. 263:F459-F465(1992).
CC   -!- FUNCTION: SODIUM-DEPENDENT GLUCOSE TRANSPORTER. HAS A NA+ TO
CC       GLUCOSE COUPLING RATIO OF 1:1.
CC   -!- FUNCTION: EFFICIENT SUBSTRATE TRANSPORT IN MAMMALIAN KIDNEY IS
CC       PROVIDED BY THE CONCERTED ACTION OF A LOW AFFINITY HIGH CAPACITY
CC       AND A HIGH AFFINITY LOW CAPACITY NA(+)/GLUCOSE COTRANSPORTER
CC       ARRANGED IN SERIES ALONG KIDNEY PROXIMAL TUBULES.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
CC   ----------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------------
DR   EMBL; M95549; AAA36608.1; -.
DR   Genew; HGNC:11037; SLC5A2.
DR   MIM; 182381; -.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   TIGRFAMs; TIGR00813; sss; 1.
DR   PROSITE; PS00456; NA_SOLUT_SYMP_1; 1.
DR   PROSITE; PS00457; NA_SOLUT_SYMP_2; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
KW   Transport; Sugar transport; Transmembrane; Sodium transport; Symport;
KW   Glycoprotein.
FT   DOMAIN       1     25       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     26    44       POTENTIAL.
FT   DOMAIN       45    61       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     62    82       POTENTIAL.
FT   DOMAIN       83    102      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     103   123      POTENTIAL.
FT   DOMAIN       124   168      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     169   188      POTENTIAL.
FT   DOMAIN       189   205      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     206   226      POTENTIAL.
FT   DOMAIN       227   270      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     271   291      POTENTIAL.
FT   DOMAIN       292   314      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     315   334      POTENTIAL.
FT   DOMAIN       335   423      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     424   443      POTENTIAL.
FT   DOMAIN       444   455      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     456   476      POTENTIAL.
```

```
FT   DOMAIN       477     526       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     527     547       POTENTIAL.
FT   DOMAIN       548     650       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     651     671       POTENTIAL.
FT   CARBOHYD     250     250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     399     399       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   SITE          40      40       IMPLICATED IN SODIUM COUPLING
FT                                  (BY SIMILARITY).
FT   SITE         300     300       IMPLICATED IN SODIUM COUPLING
FT                                  (BY SIMILARITY).
SQ   SEQUENCE   672 AA;  72896 MW;  233C65F1601B0337 CRC64;

  Query Match             9.8%;  Score 292;  DB 1;  Length 672;
  Best Local Similarity   24.1%;  Pred. No. 6.1e-12;
  Matches  147;  Conservative   91;  Mismatches  237;  Indels  136;  Gaps   22;

Qy       8 LIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY 67
           :: |  ::||:: ||:|:  || | |:          : || :     | :: |: :| |:
Db      26 ILVIAAYFLLVIGVGLWSMCRT-NRGTV----GGYFLAGRSMVWWPVGASLFASNIGSGH 80

Qy      68 INGTAEAVYVPGYGLAWAQAPIGYSLS----LILGGLFFAKPMRSKGYVTMLDPFQQIYG 123
             | |      | | |   |  |:  :    ::| |  ||    : | :||     |
Db      81 FVGLA------GTGAASGLAVAGFEWNALFVVLLLGWLFAPVYLTAGVITM----PQYLR 130

Qy     124 KRMGG------------LLFIPALMGEMFWAAAIF--SALGATISVIIDVDMHISVIISA 169
           || ||             |:|   :    :: |:|   |||  |         : |||
Db     131 KRFGGRRIRLYLSVLSLFLYIFTKISVDMFSGAVFIQQALGWNI--------YASVIALL 182

Qy     170 LIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY----- 224
             |  :||: ||| :: ||| || | |  |  | : :|         |:: :  ||
Db     183 GITMIYTVTGGLAALMYTDTVQTFVILGGACILMGYAFHEVG----GYSGLFDKYLGAAT 238

Qy     225 -----QKPWLGTVDSSEVYSWLDSFLLL---MLGGIPW-----------------QAYF 258
                : | :| : |       ||: ||   : | :||                   |
Db     239 SLTVSEDPAVGNISSFCYRPRPDSYHLLRHPVTGDLPWPALLLGLTIVSGWYWCSDQVIV 298

Qy     259 QRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKT----TE 314
           || |:  | |: :    |  :  |      :: | |      :: |  :|:        ||
Db     299 QRCLAGKSLTHIKAGCILCGYLKLTPMFLMVMPGMISRILYPDEVACVVPEVCRRVCGTE 358

Qy     315 E--ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNA 372
           :::  | ::  | |  :    |  : ||:|||  |  |:|::|  :||    |  |
Db     359 VGCSNIAYPRLVVKLMPNGLRGLMLAVMLAALMSSLASIFNSSSTLFTMDIY-TRLRPRA 417

Qy     373 SDKEIVWVMRI-TVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQLLCV----LFV 427
           |:|:: | |:  ||:   |   : ::     |  :     |   : | :  |     |||
Db     418 GDRELLLVGRLWVVFIVVVSVAWLPVVQAAQGGQLFDYIQAVSSYLAPPVSAVFVLALFV 477

Qy     428 KGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMV-- 485
             |  ||  | : || : :    |                        :| | : : |
Db     478 PRVNEQGAFWGLIGGLLMGLARLIP----------------------EFSFGSGSCVQP 514

Qy     486 ---TSFLTNICISYLAKYLFE-SGTLPPKLDVFDAVV----------ARHSEENMDKTI 530
              :||  :   | |  ||  || |   : :  | :            |||:|
Db     515 SACPAFLCGVHYLYFAIVLFFCSGLLTLTVSLCTAPIPRKHLHRLVFSLRHSKE------ 568

Qy     531 LVKNENIKLDE 541
             : |::  ||
Db     569 --EREDLDADE 577


RESULT 11
SL54_MOUSE
ID   SL54_MOUSE     STANDARD;      PRT;   656 AA.
AC   Q9ET37;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Low affinity sodium-glucose cotransporter (Sodium/glucose
DE   cotransporter 3) (Na(+)/glucose cotransporter 3).
GN   SLC5A4A.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/Sv;
RX   MEDLINE=20499361; PubMed=11042146;
RA   Pletcher M.T., Roe B.A., Chen F., Do T., Do A., Malaj E., Reeves R.H.;
RT   "Chromosome evolution: the junction of mammalian chromosomes in the
RT   formation of mouse chromosome 10.";
RL   Genome Res. 10:1463-1467(2000).
CC   -!- FUNCTION: SODIUM-DEPENDENT GLUCOSE TRANSPORTER (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
CC   ---------------------------------------------------------------------
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF251267; AAG01741.1; -.
DR   MGD; MGI:1927848; Slc5a4a.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   TIGRFAMs; TIGR00813; sss; 1.
DR   PROSITE; PS00456; NA_SOLUT_SYMP_1; FALSE_NEG.
DR   PROSITE; PS00457; NA_SOLUT_SYMP_2; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
KW   Transport; Sugar transport; Transmembrane; Sodium transport; Symport;
KW   Glycoprotein.
FT   DOMAIN         1      28       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      29      47       POTENTIAL.
FT   DOMAIN        48      64       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      65      85       POTENTIAL.
FT   DOMAIN        86     105       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     106     126       POTENTIAL.
FT   DOMAIN       127     171       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     172     191       POTENTIAL.
FT   DOMAIN       192     208       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     209     229       POTENTIAL.
FT   DOMAIN       230     270       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     271     291       POTENTIAL.
FT   DOMAIN       292     314       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     315     334       POTENTIAL.
FT   DOMAIN       335     423       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     424     443       POTENTIAL.
FT   DOMAIN       444     455       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     456     476       POTENTIAL.
FT   DOMAIN       477     526       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     527     547       POTENTIAL.
FT   DOMAIN       548     634       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     635     655       POTENTIAL.
FT   CARBOHYD     248     248       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   656 AA;  71837 MW;  A6668E815204D39B CRC64;

  Query Match             9.8%;  Score 290;  DB 1;  Length 656;
  Best Local Similarity   22.3%;  Pred. No. 8.1e-12;
  Matches  147;  Conservative   102;  Mismatches  252;  Indels  158;  Gaps   25;

Qy      11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING 70
           |:::::::: ||:||  :|  |           : || :     | :: |: :| |:  |
Db      32 IVIYFVVVMAVGVWAMLKTNRSTVG-----GFFLAGRSMTWWPMGASLFASNIGSGHFVG 86

Qy      71 TAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSK-GYVTMLDPFQQIYGKRMGG- 128
             |       |  :  |     ::| |:|| :|    |:  | | :||     :   || ||
Db      87 LAGTGAASGIAVT-AFESHSFALLLVLGWIFV--PIYIKAGVMTM----PEYLKKRFGGK 139

Qy     129 ---------LLFIPALM---GEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYT 176
                    |||  ::    ::| : |||      | : : ::::::::|   |  ::|
Db     140 RLQIYLSILFLFICVILTISADIF-SGAIF------IKLALGLNLYLAILILLAITAIFT 192
```

```
Qy    177 LVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVD--- 233
          : ||| || ||| ||   : || :| : ||     |   |: :   |:       |:
Db    193 ITGGLASVIYTDTVQAVIMLVGSFILMVFAF----VEVGGYESFTEKFMNAIPSVVEGDN 248

Qy    234 ---SSEVYS-WLDSFLLL---MLGGIPW-----------------QAYFQRVLSSSSAT 268
            :|  |:   ||| :    : | |||                 |   || |    : :
Db    249 LTINSRCYTPQPDSFHIFRDPVTGDIPWPGTAFGMPITALWYWCINQVIVQRCLCGKNLS 308

Qy    269 YAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLP---------DPKTTEEADMI 319
          : :    |  :  |:     :: | |      :  |  :|         |   |  |
Db    309 HVKAACILCGYLKLLPLFFMVMPGMISRILYTDMVACVVPSECVKHCGVDVGCTNYA--- 365

Qy    320 LPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVW 379
           |:::  | |  :    |  : |::|||  |   |||::|  ::|    |: ||::|::
Db    366 YPMLVLKLMPPGLRGLMLSVMLASLMSSLTSVFNSASTLFTIDLY-TKIRKKASERELLI 424

Qy    380 VMRITVFVFGASATAMALLTKTVYG---LWYLSSDLVYI--VIFPQLLCVLFVKGTNTYG 434
           |: | |   ::     : :   |   : |  :   |:   |    |  :| |  |  |
Db    425 AGRLFVSVLIVTSILWVPIVEVSQGGQLVHYTEAISSYLGPPIAAVFLVAVFCKRANEQG 484

Qy    435 AVAGYVSGLFL----------------------RITGGEPYLYLQPLIFYPGYYPDDNG 471
          |  | : || :                      :|  |  |||   ::|:
Db    485 AFWGLMVGLVMGLIRMIAEFSYGTGSCLAPSSCPKIICGVHYLYFAIILFF--------- 535

Qy    472 IYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDV----FDAVVARHSEENMD 527
                  |  : :|||  |        |  ||        :     || :|
Db    536 --------------VCILVILGVSYLTK---------PIPDVHLHRLCWALRNSKEERID 572

Qy    528 KTILVKNEN---IKLDELALVKPR------------QSMTLSSTFTNKEAFLDVDSSP 570
            | ||    : :|    |||              |     |   :|| |    |
Db    573 LDAEDKEENGADDRTEEDQTEKPRGCLKKTCDLFCGLQRAEFKLTKVEEEALTDTTEKP 631
```

RESULT 12
SL53_MOUSE
ID SL53_MOUSE STANDARD; PRT; 718 AA.
AC Q9JKZ2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Sodium/myo-inositol cotransporter (Na(+)/myo-inositol cotransporter).
GN SLC5A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20237552; PubMed=10773690;
RA McVeigh K.E., Mallee J.J., Lucente A., Barnoski B.L., Wu S.,
RA Berry G.T.;
RT "Murine chromosome 16 telomeric region, homologous with human
RT chromosome 21q22, contains the osmoregulatory Na(+)/myo-inositol
RT cotransporter (SLC5A3) gene.";
RL Cytogenet. Cell Genet. 88:153-158(2000).
CC -!- FUNCTION: PREVENTS INTRACELLULAR ACCUMULATION OF HIGH
CC CONCENTRATIONS OF MYO-INOSITOL (AN OSMOLYTE) THAT RESULT IN
CC IMPAIRMENT OF CELLULAR FUNCTION.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
CC ----------

CC ----------
DR EMBL; AF220915; AAF43668.1; -.
DR MGD; MGI:1858226; Slc5a3.
DR InterPro; IPR001734; Na/solut_symport.
DR Pfam; PF00474; SSF; 1.
DR TIGRFAMs; TIGR00813; sss; 1.
DR PROSITE; PS00456; NA_SOLUT_SYMP_1; 1.
DR PROSITE; PS00457; NA_SOLUT_SYMP_2; 1.
DR PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
KW Transport; Transmembrane; Sodium transport; Symport; Glycoprotein.
FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 10 29 POTENTIAL.
FT DOMAIN 30 38 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 39 57 POTENTIAL.
FT DOMAIN 58 86 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 87 110 POTENTIAL.
FT DOMAIN 111 123 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 124 144 POTENTIAL.
FT DOMAIN 145 157 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 158 183 POTENTIAL.
FT DOMAIN 184 186 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 187 205 POTENTIAL.
FT DOMAIN 206 303 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 304 324 POTENTIAL.
FT DOMAIN 325 353 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 354 376 POTENTIAL.
FT DOMAIN 377 406 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 407 430 POTENTIAL.
FT DOMAIN 431 443 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 444 462 POTENTIAL.
FT DOMAIN 463 510 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 511 532 POTENTIAL.
FT DOMAIN 533 695 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 696 716 POTENTIAL.
FT CARBOHYD 32 32 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SITE 24 24 IMPLICATED IN SODIUM COUPLING
FT (BY SIMILARITY).
FT SITE 285 285 IMPLICATED IN SODIUM COUPLING
FT (BY SIMILARITY).
SQ SEQUENCE 718 AA; 79554 MW; D035CFBECDDA803B CRC64;

Query Match 9.7%; Score 289; DB 1; Length 718;
Best Local Similarity 21.7%; Pred. No. 1e-11;
Matches 150; Conservative 113; Mismatches 209; Indels 218; Gaps 32;

```
Qy      9 IAII-VFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGG- 66
          ||:: ::::|:: :| :| |::  |  :                |  : | :|  |||  |
Db     10 IAVVALYFILVMCIGFFAMWKSNRSTVS-------------GYFLAGRSM--TWVAIGA 53

Qy     67 --YINGTAEAVYVPGYGLAWAQAPIGYS----------LSLILGGLFFAKPMRSKGYVTM 114
            :::       ::   ||| : |  |::           |  :|| :|    :|| |  ||
Db     54 SLFVSNIGSEHFI---GLAGSGAASGFAVGAWEFNALLLLQLLGWVFIPIYIRS-GVYTM 109

Qy    115 LDPFQQIYGKRMGG------------LLFIPALMGEMFWAAAIFSALGATISVIIDVDMH 162
              :   || ||                ||:|   :    :: |:|      |   :  :::
Db    110 ----PEYLSKRFGGHRIQVYFAALSLLLYIFTKLSVDLYSGALF------IQESLGWNLY 159

Qy    163 ISVIISALIATLYTLVGGLYSVAYTDVVQLFCIFVG----LWISV-----------PFAL 207
          :|||:   :  | |: ||| :| ||| :|   : :|    : ||:           : |
Db    160 VSVILLIGMTALLTVTGGLVAVIYTDTLQALLMIIGALTLMVISMVKIGGFEEVKRRYML 219

Qy    208 SHPAVADI---------GFTAVHAKYQK------------PWLGTV---DSSEVYSWLDS 243
          : | || |                || |               || | :    : |: |
Db    220 ASPDVASILLKYNLSNTNACMVHPKANALKMLRDPTDEDVPWPGFILGQTPASVWYWCAD 279

Qy    244 FLLLMLGGIPWQAYFQRVLSSSSATYAQ----VLSFLAAFGCLVMAIPAIL--------- 290
                |   ||||:: :  :|:    :  ||        :: :| ::
Db    280 ----------QVIVQRVLAAKNIAHAKGSTLMAGFLKLLPMFIIVVPGMISRIVFADEI 328

Qy    291 --------IGAIGASTDWNQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSA 342
                  :   |:    :  ||                | ::  | || :    :  : |
Db    329 ACINPEHCMQVCGSRAGCSNIAY--------------PRLVMTLVPVGLRGLMMAVMIA 373

Qy    343 AVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITV-FVFGASATAMALLTKT 401
          |:||  ||   |||::|  ::|:|  |::|| :|:: | || | |:   |   : :: :
Db    374 ALMSDLDSIFNSASTIFTLDVYKL-IRKSASSRELMIVGRIFVAFMVVISIAWVPIIVEM 432

Qy    402 VYGLWYLSSDLVYIVIFPQL----LCVLFVKGTNT----YGAVAGYVSG---LFLRITGG 450
```

```
Db    433 QGGQMYLYIQEVADYLTPPVAALFLLAIFWKRCNEQGAFYGGMAGFVLGAVRLILAFTYR 492

Qy    451 EP----------------YLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICI 494
            |                |:|:   :|:                          :|  :| : :
Db    493 APECDQPDNRPGFIKDIHYMYVATALFW---------------------ITGLIT-VIV 529

Qy    495 SYLAKYLFESGTLPPKLDVFDAVVARHSEENMDKTILV------KNENIKLDELALVK-P 547
          | |         | ||  |          | :     ||::       |:|  |: | ::::
Db    530 SLL--------TPPPTKDQI-----RTTTFWSKKTLVTKESCSQKDEPYKMQEKSILQCS 576

Qy    548 RQSMTLSSTFTNKEAFLDVDSSPEGSGTED 577
            |  :| |  | ::    : | :|   ||
Db    577 ENSEVISHTIPNGKS----EDSIKGLQPED 602


RESULT 13
OPUE_BACSU
ID   OPUE_BACSU     STANDARD;      PRT;   492 AA.
AC   O06493;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Osmoregulated proline transporter (Sodium/proline symporter).
GN   OPUE.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / JH642;
RA   von Blohn C., Kempf B., Kappes R.M., Bremer E.;
RT   "Osmostress response in Bacillus subtilis: characterization of a
RT   proline uptake system (OpuE) regulated by high osmolarity and the
RT   alternative transcription factor sigma B.";
RL   Mol. Microbiol. 25:175-187(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=97124186; PubMed=8969499;
RA   Borriss R., Porwollik S., Schroeter R.;
RT   "The 52 degrees-55 degrees segment of the Bacillus subtilis
RT   chromosome: a region devoted to purine uptake and metabolism, and
RT   containing the genes cotA, gabP and guaA and the pur gene cluster
RT   within a 34960 bp nucleotide sequence.";
RL   Microbiology 142:3027-3031(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- FUNCTION: CATALYZES THE SODIUM-DEPENDENT UPTAKE OF EXTRACELLULAR
CC       PROLINE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
CC   ---------------------------------------------------------------------
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U92466; AAB66512.1; -.
DR   EMBL; AF011545; AAB72182.1; -.
DR   EMBL; Z99107; CAB12486.1; -.
DR   SubtiList; BG12641; opuE.
DR   InterPro; IPR002259; DER/eqnu_transpt.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   PRINTS; PR01130; DERENTRNSPRT.
DR   TIGRFAMs; TIGR00813; sss; 1.
DR   PROSITE; PS00456; NA_SOLUT_SYMP_1; 1.
DR   PROSITE; PS00457; NA_SOLUT_SYMP_2; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
KW   Transport; Transmembrane; Sodium transport; Symport;
KW   Complete proteome.
FT   TRANSMEM      3     23       POTENTIAL.
FT   TRANSMEM     62     82       POTENTIAL.
FT   TRANSMEM    125    145       POTENTIAL.
FT   TRANSMEM    161    181       POTENTIAL.
FT   TRANSMEM    191    211       POTENTIAL.
FT   TRANSMEM    224    244       POTENTIAL.
FT   TRANSMEM    271    291       POTENTIAL.
FT   TRANSMEM    314    334       POTENTIAL.
FT   TRANSMEM    365    385       POTENTIAL.
FT   TRANSMEM    394    414       POTENTIAL.
FT   TRANSMEM    424    444       POTENTIAL.
FT   TRANSMEM    449    469       POTENTIAL.
SQ   SEQUENCE   492 AA;  53282 MW;  23459873F1E799E6 CRC64;

  Query Match             9.6%;  Score 285;  DB 1;  Length 492;
  Best Local Similarity   22.1%;  Pred. No. 1.2e-11;
  Matches  118;  Conservative   97;  Mismatches  214;  Indels  106;  Gaps   18;

Qy       5 VEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVG 64
            :| :|:: :::: :||:| :|  :| :              ::||| :|  |   :  | :
Db       3 IEIIISLGIYFIAMLLIGWYAFKKTTDIND-------YMLGGRGLGPFVTALSAGAADMS 55

Qy      65 GGYINGTAEAVYVPGYGLAWAQAPI---GYSLSLILGGLFFAKPMRSKGYVTMLDPFQQI 121
           |  : |   |:: |   |    |    :    ||  |:|    |    :   :|: | | :
Db      56 GWMLMGVPGAMFATGLSTLWLALGLTIGAYSNYLLLAPRLRAYTEAADDAITIPDFFDKR 115

Qy     122 YGKRMGGLLFIPALMGEMFWAAAIFSAL---GATISVIIDVDMHISVIISALIATLYTLV 178
           :     .| : ||:  :|:     | :  |         |  : : ::  :  ||||
Db     116 FQHSSSLLKIVSALIIMIFFTLYTSSGMVSGGRLFESAFGADYKLGLFLTTAVVVLYTLF 175

Qy     179 GGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVY 238
           ||  :|: || ||   :|  | : ||          : ||  |     |  :|:   :
Db     176 GGFLAVSLTDFVQGAIMFAAL-VLVPI---------VAFT--HVGGVAPTFHEIDAVNPH 223

Qy     239 SWLDSF----LLLMLGGIPWQAYFQRVLSSSSATYAQ---VLSFLA------------- 277
             || |    :: ::  : |  :                 | |   :: |:|
Db     224 -LLDIFKGASVISIISYLAWGLGY----------YGQPHIIVRFMAIKDIKDLKPARRIG 272
```

```
Qy     278 AFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGL 337
Db     273 MSWMIITVLGSVLTGLIG-------VAYAHKFGVAVKDPEMIFIIFSKILFHPLITGFLL 325

Qy     338 GAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMAL 397
Db     326 SAILAAIMSSISSQLLVTASAVTEDLYRSFFRRKASDKELVMIGRLSVLVIAVIAVLLSL 385

Qy     398 LTKTVYGLWYLSSDLVYIVIF----------PQLLCVLFVKGTNTYGAVAGYVSG---LF 444
Db     386 NP---------NSTILDLVGYAWAGFGSAFGPAILLSLYWKRMNEWGALAAMIVGAATVL 436

Qy     445 LRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAK 499
Db     437 IWITTG--------------LAKSTGVY-EIIP----GFILSMIAGIIVSMITK 471
```

RESULT 14
SL53_CANFA
ID SL53_CANFA STANDARD; PRT; 718 AA.
AC P31637;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Sodium/myo-inositol cotransporter (Na(+)/myo-inositol cotransporter).
GN SLC5A3 OR SMIT.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=92210609; PubMed=1372904;
RA Kwon H.M., Yamauchi A., Uchida S., Preston A.S., Garcia-Perez A.,
RA Burg M.B., Handler J.S.;
RT "Cloning of the cDNA for a Na+/myo-inositol cotransporter, a
RT hypertonicity stress protein.";
RL J. Biol. Chem. 267:6297-6301(1992).
CC -!- FUNCTION: PREVENTS INTRACELLULAR ACCUMULATION OF HIGH
CC CONCENTRATIONS OF MYO-INOSITOL (AN OSMOLYTE) THAT RESULT IN
CC IMPAIRMENT OF CELLULAR FUNCTION.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: BRAIN AND KIDNEY.
CC -!- INDUCTION: MEDIUM HYPERTONICITY.
CC -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; M85068; -; NOT_ANNOTATED_CDS.
DR PIR; A42163; A42163.
DR InterPro; IPR001734; Na/solut_symport.
DR Pfam; PF00474; SSF; 1.
DR TIGRFAMs; TIGR00813; sss; 1.
DR PROSITE; PS00456; NA_SOLUT_SYMP_1; 1.
DR PROSITE; PS00457; NA_SOLUT_SYMP_2; 1.
DR PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
KW Transport; Transmembrane; Sodium transport; Symport; Glycoprotein.

FT	DOMAIN	1	9	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	10	29	POTENTIAL.
FT	DOMAIN	30	38	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	39	57	POTENTIAL.
FT	DOMAIN	58	86	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	87	110	POTENTIAL.
FT	DOMAIN	111	123	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	124	144	POTENTIAL.
FT	DOMAIN	145	157	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	158	183	POTENTIAL.
FT	DOMAIN	184	186	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	187	205	POTENTIAL.
FT	DOMAIN	206	303	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	304	324	POTENTIAL.
FT	DOMAIN	325	353	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	354	376	POTENTIAL.
FT	DOMAIN	377	406	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	407	430	POTENTIAL.
FT	DOMAIN	431	443	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	444	462	POTENTIAL.
FT	DOMAIN	463	510	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	511	532	POTENTIAL.
FT	DOMAIN	533	695	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	696	716	POTENTIAL.
FT	CARBOHYD	32	32	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SITE	24	24	IMPLICATED IN SODIUM COUPLING (BY SIMILARITY).
FT	SITE	285	285	IMPLICATED IN SODIUM COUPLING (BY SIMILARITY).

SQ SEQUENCE 718 AA; 79545 MW; 4C1B5CC4485CD268 CRC64;

Query Match 9.4%; Score 278.5; DB 1; Length 718;
Best Local Similarity 23.1%; Pred. No. 5e-11;
Matches 154; Conservative 115; Mismatches 222; Indels 177; Gaps 34;

```
Qy       9 IAII-VFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGG- 66
Db      10 IAIVALYFILVMCIGFFAMWKSNRSTVS-------------GYFLAGRSM--TWVAIGA 53

Qy      67 --YINGTAEAVYVPGYGLAWAQAPIGYS----------LSLILGGLFFAKPMRSKGYVTM 114
Db      54 SLFVSNIGSEHFI---GLAGSGAASGFAVGAWEFNALLLLQLLGWVFIPIYIRS-GVYTM 109

Qy     115 LDPFQQIYGKRMGG-----------LLFIPALMGEMFWAAAIFSALGATISVIIDVDMH 162
Db     110 ----PEYLSKRFGGHRIQVYFAALSLILYIFTKLSVDLYSGALF------IQESLGWNLY 159

Qy     163 ISVIISALIATLYTLVGGLYSVAYTDVVQLFCIFVG----LWISV----------PFAL 207
Db     160 VSVILLIGMTALLTVTGGLVAVIYTDTLQALLMIVGALTLMIISMMEIGGFEEVKRRYML 219

Qy     208 SHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSWLDSFLLL----------MLGGIP- 253
Db     220 ASPNVTSILLT-----YN---LSNTNSCNVHPKKDALKMLRNPTDEDVPWPGFVLGQTPA 271

Qy     254 ----W---QAYFQRVLSSSSATYAQ----VLSFLAAFGCLVMAIPAIL---------- 290
Db     272 SVWYWCADQVIVQRVLAAKNIAHAKGSTLMAGFLKLLPMFIIVVPGMISRILFADDIACI 331

Qy     291 -----IGAIGASTDWNQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVM 345
Db     332 NPEHCMQVCGSRAGCSNIAY----------PRLVMKLVPVGLRGLMMAVMIAALM 376

Qy     346 SSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITV-FVFGASATAMALLTKTVYG 404
Db     377 SDLDSIFNSASTIFTLDVYKL-IRRSASSRELMIVGRIFVAFMVVISIAWVPIIVEMQGG 435

Qy     405 LWYLSSDLVYIVIFPQL----LCVLFVKGTNT----YGAVAGYVSGLFLRITGGEPYLYL 456
Db     436 QMYLYIQEVADYLTPPVAALFLLAIFWKRCNEQGAFYGGMAGFVLGA-VRLT--LAFAYR 492

Qy     457 QPLIFY----PGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLD 512
Db     493 APECDQPDNRPGFIKDIHYMY----VATALFWVTGLIT-VIVSLL--------TPPPTKE 539

Qy     513 VFDAVVARHSEENMDKTILVKNENIKLDELALVKPRQSMTLSSTFTNKEAFLDVDSSPEG 572
Db     540 QI-----RTTTFWSKKSLVVKESCSPKDEPYKMQEKSILRCSE---NSEATNHI--IPNG 589

Qy     573 SGTEDNLQ 580
Db     590 K-SEDSIK 596
```

```
RESULT 15
SL53_BOVIN
ID   SL53_BOVIN     STANDARD;      PRT;   718 AA.
AC   P53793;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Sodium/myo-inositol cotransporter (Na(+)/myo-inositol cotransporter).
GN   SLC5A3 OR SMIT.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mallee J.J., Parrella T., Kwon H.M., Berry G.T.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PREVENTS INTRACELLULAR ACCUMULATION OF HIGH
CC       CONCENTRATIONS OF MYO-INOSITOL (AN OSMOLYTE) THAT RESULT IN
CC       IMPAIRMENT OF CELLULAR FUNCTION.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U41338; AAA93188.1; -.
DR   InterPro; IPR001734; Na/solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   TIGRFAMs; TIGR00813; sss; 1.
DR   PROSITE; PS00456; NA_SOLUT_SYMP_1; 1.
DR   PROSITE; PS00457; NA_SOLUT_SYMP_2; 1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
KW   Transport; Transmembrane; Sodium transport; Symport; Glycoprotein.
FT   DOMAIN        1      9       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      10     29      POTENTIAL.
FT   DOMAIN        30     38      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      39     57      POTENTIAL.
FT   DOMAIN        58     86      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      87     110     POTENTIAL.
FT   DOMAIN        111    123     EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      124    144     POTENTIAL.
FT   DOMAIN        145    157     CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      158    183     POTENTIAL.
FT   DOMAIN        184    186     EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      187    205     POTENTIAL.
FT   DOMAIN        206    303     CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      304    324     POTENTIAL.
FT   DOMAIN        325    353     EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      354    376     POTENTIAL.
FT   DOMAIN        377    406     CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      407    430     POTENTIAL.
FT   DOMAIN        431    443     EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      444    462     POTENTIAL.
FT   DOMAIN        463    510     CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      511    532     POTENTIAL.
FT   DOMAIN        533    695     CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      696    716     POTENTIAL.
FT   CARBOHYD      32     32      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   SITE          24     24      IMPLICATED IN SODIUM COUPLING
FT                                (BY SIMILARITY).
FT   SITE          285    285     IMPLICATED IN SODIUM COUPLING
FT                                (BY SIMILARITY).
SQ   SEQUENCE   718 AA;  79673 MW;  206BE25FA385111D CRC64;
```

```
Query Match              9.3%;  Score 275;  DB 1;  Length 718;
Best Local Similarity   22.2%;  Pred. No. 8.5e-11;
Matches  148;  Conservative  122;  Mismatches  225;  Indels  172;  Gaps   32;

Qy       9 IAII-VFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGG- 66
           ||| : :::|:: :| :| |::  |  :                 |  : | :|  ||| |
Db      10 IAIVALYFILVMCIGFFAMWKSNRSTVS-------------GYFLAGRSM--TWVAIGA 53

Qy      67 --YINGTAEAVYVPGYGLAWAQAPIGYS----------LSLILGGLFFAKPMRSKGYVTM 114
             :::      ::   ||| : |  |::             |  :|| :|     :|| |  ||
Db      54 SLFVSNIGSEHFI---GLAGSGAASGFAVGAWEFNALLLLQLLGWVFIPIYIRS-GVYTM 109

Qy     115 LDPFQQIYGKRMGG-----------LLFIPALMGEMFWAAAIFSALGATISVIIDVDMH 162
             :  || ||           :|:|   :    :: |:|      |  :  :::
Db     110 ----PEYLSKRFGGHRIQVYFAALSLILYIFTKLSVDLYSGALF------IQESMGWNLY 159

Qy     163 ISVIISALIATLYTLVGGLYSVAYTDVVQLFCIFVG----LWISV----------PFAL 207
           :|||:   :  | |: ||| :| ||| :|   : ||    : ||:           : |
Db     160 VSVILLIGMTALLTVTGGLVAVIYTDTLQALLMIVGALTLMVISMMEIGGFEEVKRRYML 219

Qy     208 SHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSWLDSFLLL----------MLGGIP- 253
           : | |  |  |     |    :|  |:   |:  :|                :|| |
Db     220 ASPNVTSILLT-----YN---LSNTNSCNVHPKKDALKMLRNPTDEDVPWPGFILGQTPA 271

Qy     254 ----W---QAYFQRVLSSSSATYAQ----VLSFLAAFGCLVMAIPAIL---------- 290
               |   |  ||||:: :  :|:    :  ||      :: :| ::
Db     272 SVWYWCADQVIVQRVLAAKNIAHAKGSTLMAGFLKLLPMFIIVVPGMISRILFADDIACI 331

Qy     291 -----IGAIGASTDWNQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVM 345
                :  |:    :  ||                | ::  | || :   :  : ||:|
Db     332 NPEHCMQVCGSRAGCSNIAY----------PRLVMKLVPVGLRGLMMAVMIAALM 376

Qy     346 SSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITV-FVFGASATAMALLTKTVYG 404
           |  ||   |||::|  ::|:|  |::|| :|:: | || | |:   |   : :: :   |
Db     377 SDLDSIFNSASTIFTLDVYKL-IRKSASSRELMIVGRIFVAFMVVISIAWVPIIVEMQGG 435

Qy     405 LWYLSSDLVYIVIFPQL----LCVLFVKGTNTYGAVAGYVSGLFL---RITGGEPYLYLQ 457
             ||    |   : | :    |  :| |  |  || ||:|  |   |:|    : |
Db     436 QMYLYIQEVADYLTPPVAALFLLAIFWKRCNEQGAFYGGMAGFILVVVRLT--LAFAYRA 493

Qy     458 PLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAV 517
           |          ||:  :: :   :  :|    ::| : |: :   |    | ||  :
Db     494 P----ECDQPDNRPVFIKDIHYMYVATALFWITGL-ITVIVSLL----TPPPTKEQI--- 541

Qy     518 VARHSEENMDKTILV------KNENIKLDELALVK-PRQSMTLSSTFTNKEAFLDVDSSP 570
           | :     |:::|      |:|  |: | ::::    |  ::    | ::    : |
Db     542 --RTTTFWSKKSLVVKESCSPKDEPYKMQEKSILRCSENSEVINHVIPNGKS----EDSI 595

Qy     571 EGSGTED 577
           :|   ||
Db     596 KGLQPED 602


Search completed: March 31, 2003, 18:35:24
Job time : 42 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: March 31, 2003, 18:28:26 ; Search time 49 Seconds
(without alignments)
1137.918 Million cell updates/sec

Title: US-10-069-541-6
Perfect score: 2972
Sequence: 1 MAFHVEGLIAIIVFYLLILL.....EAFLDVDSSPEGSGTEDNLQ 580

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2972	100.0	580	2	JC7502	choline transporte
2	1361.5	45.8	631	2	T20037	hypothetical prote
3	344	11.6	492	2	D75188	proline symporter
4	308.5	10.4	662	2	A37226	glucose transport
5	306	10.3	664	2	A33545	Na+/glucose cotran
6	306	10.3	665	2	A53582	Na+/glucose cotran
7	303.5	10.2	660	2	A44432	amino acid transpo
8	301	10.1	463	2	E83468	probable sodium/so
9	299.5	10.1	507	2	B83988	proline transporte
10	298	10.0	672	2	A42251	nucleoside transpo
11	294	9.9	664	2	S59637	glucose transport
12	292	9.8	672	2	A56765	sodium-glucose cot
13	288	9.7	664	2	S59638	glucose transport
14	286	9.6	491	2	H71097	hypothetical prote
15	285	9.6	492	2	H69670	sodium/proline sym
16	284	9.6	537	2	C71008	probable proline p
17	283	9.5	484	2	E75138	osmoregulated prol
18	282.5	9.5	501	2	T44298	sodium/proline sym
19	279.5	9.4	496	2	H82382	sodium/proline sym
20	278.5	9.4	718	2	A42163	Na+/myo-inositol c
21	271	9.1	537	2	A75123	proline permease (
22	270.5	9.1	718	2	A56851	Na+/myo-inositol c
23	267	9.0	522	2	B84247	proline permease [
24	263.5	8.9	494	2	JC2382	sodium/proline sym
25	261	8.8	512	2	E89978	high affinity prol
26	260.5	8.8	504	2	E64118	sodium/proline sym
27	260.5	8.8	605	2	A36361	glucose transport
28	259	8.7	526	2	C69115	sodium/proline sym
29	254	8.5	496	2	A71980	sodium/proline sym
30	254	8.5	502	2	E90786	major sodium/proli
31	254	8.5	502	2	C85646	major sodium/proli
32	253.5	8.5	497	2	T48676	proline uptake pro
33	253	8.5	502	1	JGECPP	sodium/proline sym
34	252	8.5	461	2	G83610	probable sodium/so
35	249	8.4	496	2	G64526	sodium/proline sym
36	249	8.4	502	2	AF0633	sodium/proline sym
37	246.5	8.3	512	2	S75887	hypothetical prote
38	246	8.3	502	2	S10220	sodium/proline sym
39	243.5	8.2	506	2	F83547	sodium/proline sym
40	243	8.2	497	2	AH2015	sodium/solute symp
41	242	8.1	449	2	B69759	sodium/proline sym
42	239.5	8.1	484	2	H64105	pantothenate trans
43	239	8.0	549	2	B65215	hypothetical 59.2
44	237.5	8.0	577	2	T28017	hypothetical prote
45	236	7.9	491	2	E69383	pantothenate perme

ALIGNMENTS

RESULT 1
JC7502
choline transporter - human
C;Species: Homo sapiens (man)
C;Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 01-Dec-2000
C;Accession: JC7502
R;Apparsundaram, S.; Ferguson, S.M.; George Jr., A.L.; Blakely, R.D.
Biochem. Biophys. Res. Commun. 276, 862-867, 2000
A;Title: Molecular cloning of a human, hemicholinium-3-sensitive choline transporte
A;Reference number: JC7502
A;Contents: Spinal cord
A;Accession: JC7502
A;Molecule type: mRNA
A;Residues: 1-580 <APP>
A;Cross-references: GB:AF276871
C;Comment: This protein, a hemicholinium-3-sensitive phosphorylated transmembrane p
C;Genetics:
A;Gene: cht
A;Map position: 2q12
C;Keywords: choline transport; spinal cord; transmembrane protein; transport protei

Query Match 100.0%; Score 2972; DB 2; Length 580;
Best Local Similarity 100.0%; Pred. No. 8.2e-211;
Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60

Qy      61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120

Qy     121 IYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 IYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGG 180

Qy     181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSW 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSW 240

Qy     241 LDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDW 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 LDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDW 300

Qy     301 NQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 NQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFA 360

Qy     361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 RNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQ 420
```

```
Qy    421 LLCVLFVKGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480
Db    421 LLCVLFVKGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFK 480

Qy    481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAVVARHSEENMDKTILVKNENIKLD 540
Db    481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAVVARHSEENMDKTILVKNENIKLD 540

Qy    541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
Db    541 ELALVKPRQSMTLSSTFTNKEAFLDVDSSPEGSGTEDNLQ 580
```

RESULT 2
T20037
hypothetical protein C48D1.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T20037
R;Burton, J.
Submitted to the EMBL Data Library, October 1996
A;Reference number: Z19214
A;Accession: T20037
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-631 <WIL>
A;Cross-references: EMBL:Z81049; PIDN:CAB02847.1; GSPDB:GN00022; CESP:C48D1.3
A;Experimental source: clone C48D1
C;Genetics:
A;Gene: CESP:C48D1.3
A;Map position: 4
A;Introns: 82/1; 120/3; 187/1; 236/3; 249/1; 358/1; 510/3; 570/2

Query Match 45.8%; Score 1361.5; DB 2; Length 631;
Best Local Similarity 46.8%; Pred. No. 2.3e-92;
Matches 290; Conservative 91; Mismatches 146; Indels 93; Gaps 10;

```
Qy      7 GLIAIIVFYLLILLVGIWAAWRTKNSGSAEER----SEAIIVGGRDIGLLVGGFTMTATW 62
Db      6 GIVAIVFFYVLILVVGIWAGRKSKSSKELESEAGAATEEVMLAGRNIGTLVGIFTMTATW 65

Qy     63 VGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDP----- 117
Db     66 VGGAYINGTAEALY--NGGLLGCQAPVGYAISLVMGGLLFAKKMREEGYITMLDPFQFWN 123

Qy    118 --------------------------------FQQIYGKRMGGLLFIPALMGEMFWAAAIF 146
Db    124 FLELIFGRTFDNFRKLGRFLKLQTIIEILDFFQHKYGQRIGGLMYVPALLGETFWTAAIL 183

Qy    147 SALGATISVIIDVDMHISVIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFA 206
Db    184 SALGATLSVILGIDMNASVTLSACIAVFYTFTGGYYAVAYTDVVQLFCIFVGLLILGLYV 243

Qy    207 LSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSWLDSFLLLMLGGIPWQAYFQRVLSSSS 266
Db    244 QNRPN--------------------RFKETSLWIDCMLLLVFGGIPWQVYFQRVLSSKT 282

Qy    267 ATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKTTEEA------DMIL 320
Db    283 AHGAQTLSFVAGVGCILMAIPPALIGAIARNTDWRMTDYSPWNNGTKVESIPPDKRNMVV 342

Qy    321 PIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWV 380
Db    343 PLVFQYLTPRWVAFIGLGAVSAAVMSSADSSVLSAASMFAHNIWKLTIRPHASEKEVIIV 402

Qy    381 MRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQLLCVLFVKGTNTYGAVAGYV 440
Db    403 MRIAIICVGIMATIMALTIQSIYGLWYLCADLVYVILFPQLLCVVYMPRSNTYGSLAGYA 462

Qy    441 SGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKY 500
Db    463 VGLVLRLIGGEPLVSLPAFFHYPMY---TDGV--QYFPFRTTAMLSSMATIYIVSIQSEK 517

Qy    501 LFESGTLPPKLDVFDAVVARHSEENMDKTILVKNENIKLDELALVKPRQSMTLSSTFTNK 560
Db    518 LFKSGRLSPEWDVMGCVV----------------NIPIDHVPLPSD-VSFAVSSETLNM 559

Qy    561 EAFLDVDSSPEGSGTEDNLQ 580
Db    560 KVECDGMQFPQ-LQTEHRLQ 578
```

RESULT 3
D75188
proline symporter (proline permease). PAB2354 - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: D75188
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome
A;Reference number: A75001
A;Accession: D75188
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-492 <KAW>
A;Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB48955.1; PID:g54
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: putP-3; PAB2354
C;Superfamily: proline carrier protein

Query Match 11.6%; Score 344; DB 2; Length 492;
Best Local Similarity 24.2%; Pred. No. 1.1e-17;
Matches 132; Conservative 99; Mismatches 196; Indels 118; Gaps 25;

```
Qy      8 LIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY 67
Db     14 LVAFLFTLILPILVGFYAMKRTKS-------EEDFFVGGRAMDKITVALSAVSSGRSSWL 66

Qy     68 INGTAEAVYVPGYGLAWAQAPIGYSLS----LILGGLFFAKPMRSKGYVTMLDPFQQIYG 123
Db     67 VLGLSGMAYKMGVTAVW--AAVGYIVAEMFQFVYMGIRLRKFSERFNAITVPDYFEARFR 124

Qy    124 K-----RMGG----LLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATL 174
Db    125 DTSKILRIAASIIIIIFLTSYVGAQFNAGA------KTLSTALGISIFTALMISVLMIIV 178

Qy    175 YTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFT--------AVHAKYQK 226
Db    179 YMILGGFIAVAYNDVIRAVIMIIGLVV-----LPVIAVAKVGGTEEVLKVLHALDPKLIN 233

Qy    227 PW---LGTVDSSEVYSWLDSFLLLMLG-GIPWQAY-FQRVLSSSSATYAQVLSFLAAFGC 281
Db    234 PWAFGAGVVIG-----------FLGIGFGSPGQPHIIVRYMSIDDPNKLRVSTVVGTFWN 282

Qy    282 LVMAIPAILIGAIGASTDWNQTAYGLPDPKTT--EEADMILP-IVLQYLCPVYISFFGLG 338
Db    283 VVLAWGAIFVGLAGRAI--------VPDVSQLPGKNAEMIYPYLSAQYFPPILYGIL-IG 333

Qy    339 AVSAAVMSSADSSILSASSMFARNIYQLSFRQNA--SDKEIVWVMRITVFVFGASATAMA 396
Db    334 GIFAAILSTADSQLLVVASTVVKDLYQEVIKKGTKIDEKTALTISRVTVLVVGFLAAILA 393

Qy    397 LLTKTVYGLWYLSSDLVY-IVIF----------PQLLCVLFVKGTNTYGAVAGYVSGLFL 445
Db    394 ----------YVAKDIIFWFVLFAWGGLGASFGPTLILSLYWKGTTKWGVLAGMIVGTIT 443

Qy    446 RITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESG 505
Db    444 TIVW---KLYLKPI----------TGLY-ELVP----AFIFSLIATIIVSMITK------ 479

Qy    506 TLPPK 510
Db    480 --PPE 482
```

RESULT 4
A37226
glucose transport protein - rabbit
N;Alternate names: sodium/D-glucose cotransporter
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 30-Dec-1991 #sequence_revision 01-Mar-1996 #text_change 20-Aug-1999
C;Accession: S00515; S15974; A37226
R;Hediger, M.A.; Coady, M.J.; Ikeda, T.S.; Wright, E.M.
Nature 330, 379-381, 1987
A;Title: Expression cloning and cDNA sequencing of the Na/glucose co-transporter.
A;Reference number: S00515; MUID:88065856; PMID:2446136
A;Accession: S00515
A;Molecule type: mRNA
A;Residues: 1-662 <HED>
A;Cross-references: EMBL:X06419; NID:g1640; PIDN:CAA29727.1; PID:g1641
R;Morrison, A.I.; Panayotova-Heiermann, M.; Feigl, G.; Schoelermann, B.; Kinne, R.K.H.
Biochim. Biophys. Acta 1089, 121-123, 1991
A;Title: Sequence comparison of the sodium-D-glucose cotransport systems in rabbit renal
A;Reference number: S15974; MUID:91223090; PMID:2025641
A;Accession: S15974
A;Molecule type: mRNA
A;Residues: 1-662 <MOR>
A;Cross-references: EMBL:X55355; NID:g1716; PIDN:CAA39040.1; PID:g1717
R;Coady, M.J.; Pajor, A.M.; Wright, E.M.
Am. J. Physiol. 259, C605-C610, 1990
A;Title: Sequence homologies among intestinal and renal Na(+)/glucose cotransporters.
A;Reference number: A37226; MUID:91023017; PMID:2221040
A;Accession: A37226
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 178-662 <COA>
A;Cross-references: GB:X06419
A;Experimental source: renal cortex
C;Superfamily: proline carrier protein

Query Match 10.4%; Score 308.5; DB 2; Length 662;
Best Local Similarity 23.4%; Pred. No. 6.5e-15;
Matches 154; Conservative 110; Mismatches 238; Indels 155; Gaps 26;

```
Qy     11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING 70
Db     32 IVIYFLVVMAVGLWAMFST-NRGTV----GGFFLAGRSMVWWPIGASLFASNIGSGHFVG 86

Qy     71 TAEAVYVPGYGLAWAQAPIGYS-----LSLILGGLFFAKPMRSKGYVTMLDPFQQIY-GK 124
Db     87 LA------GTGAASGIATGGFEWNALIMVVVLGWVFVPIYIRA-GVVTMPEYLQKRFGGK 139

Qy    125 RMGGLLFIPALMGEMFW--AAAIFSALGAT-ISVIIDVDMHISVIISALIATLYTLVGGL 181
Db    140 RIQIYLSILSLLLYIFTKISADIFS--GAIFIQLTLGLDIYVAIIILLVITGLYTITGGL 197

Qy    182 YSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY----------------Q 225
Db    198 AAVIYTDTLQTAIMMVGSVILTGFAFHEVG----GYEAFTEKYMRAIPSQISYGNTSIPQ 253

Qy    226 KPWLGTVDSSEVYSWLDSFLLLMLGGIPW------------------QAYFQRVLSSSSA 267
Db    254 KCYTPREDAFHI------FRDAITGDIPWPGLVFGMSILTLWYWCTDQVIVQRCLSAKNL 307

Qy    268 TYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDP------KTTEEADMILP 321
Db    308 SHVKAGCILCGYLKVMPMFLIVMMGMVSRILYTDKVACVVPSECERYCGTRVGCTNIAFP 367

Qy    322 IVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVM 381
Db    368 TLVVELMPNGLRGLMLSVMMASLMSSLTSIFNSASTLFTMDIY-TKIRKKASEKELMIAG 426

Qy    382 RI-TVFVFGASATAMALLTKTVYG--LWYLSSDLVYI--VIFPQLLCVLFVKGTNTYGAV 436
Db    427 RLFMLFLIGISIAWVPIVQSAQSGQLFDYIQSITSYLGPPIAAVFLLAIFWKRVNEPGAF 486
```

```
Qy    437 AGYVSGLFLRI----------TG------------GEPYLYLQPLIFYPGYYPDDNGIY 473
Db    487 WGLVLGFLIGISRMITEFAYGTGSCMEPSNCPTIICGVHYLYFAIILF------------ 534

Qy    474 NQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAVVA-RHSEENMDKTILV 532
Db    535 ----------VISIITVVVVSLFTKPI-------PDVHLYRLCWSLRNSKE-------- 568

Qy    533 KNENIKLD--ELALVKPRQSMTLSSTFTNKEAF----------LDVDSSPEGSGTED 577
Db    569 --ERIDLDAGEEDIQEAPEEATDTEVPKKKKGFFRRAYDLFCGLDQDKGPKMTKEEE 623
```

RESULT 5
A33545
Na+/glucose cotransporter SGLT1 - human
C;Species: Homo sapiens (man)
C;Date: 27-Feb-1990 #sequence_revision 27-Feb-1990 #text_change 20-Aug-1999
C;Accession: A33545; A53804
R;Hediger, M.A.; Turk, E.; Wright, E.M.
Proc. Natl. Acad. Sci. U.S.A. 86, 5748-5752, 1989
A;Title: Homology of the human intestinal Na(+)/glucose and Escherichia coli Na(+)/p
A;Reference number: A33545; MUID:89345544; PMID:2490366
A;Accession: A33545
A;Molecule type: mRNA
A;Residues: 1-664 <HED>
A;Cross-references: GB:M24847; NID:g338054; PIDN:AAA60320.1; PID:g338055
R;Turk, E.; Martin, M.G.; Wright, E.M.
J. Biol. Chem. 269, 15204-15209, 1994
A;Title: Structure of the human Na+/glucose cotransporter gene SGLT1.
A;Reference number: A53804; MUID:94253082; PMID:8195156
A;Accession: A53804
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-45 <TUR>
A;Note: sequence extracted from NCBI backbone (NCBIN:147993, NCBIP:147994)
C;Genetics:
A;Gene: GDB:SLC5A1; SGLT1
A;Cross-references: GDB:120375; OMIM:182380
A;Map position: 22q13.1-22q13.1
C;Superfamily: proline carrier protein
C;Keywords: transmembrane protein; transport protein

Query Match 10.3%; Score 306; DB 2; Length 664;
Best Local Similarity 22.8%; Pred. No. 9.9e-15;
Matches 148; Conservative 104; Mismatches 218; Indels 178; Gaps 30;

```
Qy     11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING 70
Db     32 IVIYFVVVMAVGLWAMFST-NRGTV----GGFFLAGRSMVWWPIGASLFASNIGSGHFVG 86

Qy     71 TAEAVYVPGYGLAWAQAPIGYS-----LSLILGGLFFAKPMRSK-GYVTMLDPFQQIYGK 124
Db     87 LA------GTGAASGIAIGGFEWNALVLVVVLGWLFV--PIYIKAGVVTM----PEYLRK 134

Qy    125 RMGG------------LLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIA 172
Db    135 RFGGQRIQVYLSLLSLLLYIFTKISADIFSGAIF------INLALGLNLYLAIFLLLAIT 188

Qy    173 TLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQK--PWL- 229
Db    189 ALYTITGGLAAVIYTDTLQTVIMLVGSLILTGFAFHEVG----GYDAFMEKYMKAIPTIV 244

Qy    230 ---GTVDSSEVYS-WLDSFLLL---MLGGIPW------------------QAYFQRVLSS 264
Db    245 SDGNTTFQEKCYTPRADSFHIFRDPLTGDLPWPGFIFGMSILTLWYWCTDQVIVQRCLSA 304

Qy    265 SSATYAQ----VLSFLAAFGCLVMAIPAIL--------IGAI---------GASTDWNQT 303
Db    305 KNMSHVKGGCILCGYLKLMPMFIMVMPGMISRILYTEKIACVVPSECEKYCGTKVGCTNI 364

Qy    304 AYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNI 363
```

```
Db    365 AY--------------PTLVVELMPNGLRGLMLSVMLASLMSSLTSIFNSASTLFTMDI 409

Qy    364 YQLSFRQNASDKEIVWVMRITVFV-FGASATAMALLTKTVYG--LWYLSSDLVYI--VIF 418
Db    410 Y-AKVRKRASEKELMIAGRLFILVLIGISIAWVPIVQSAQSGQLFDYIQSITSYLGPPIA 468

Qy    419 PQLLCVLFVKGTNTYGAVAGYVSGLFLRI----------TG------------GEPYLY 455
Db    469 AVFLLAIFWKRVNEPGAFWGLILGLLIGISRMITEFAYGTGSCMEPSNCPTIICGVHYLY 528

Qy    456 LQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFD 515
Db    529 FAIILF----------------------AISFITIVVISLLTKPI-------PDVHLYR 558

Qy    516 AV--VARHSEENMDKTILVKNENIKLDELALVKPRQSMTLSSTFTNKE 561
Db    559 LCWSLRNSKEERID--LDAEEENIQ------EGPKETIEIETQVPEKK 598


RESULT 6
53582
a+/glucose cotransporter SGLT1 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 12-Apr-1995 #sequence_revision 12-Apr-1995 #text_change 20-Aug-1999
C;Accession: A53582
R;Lee, W.S.; Kanai, Y.; Wells, R.G.; Hediger, M.A.
J. Biol. Chem. 269, 12032-12039, 1994
A;Title: The high affinity Na(+)/glucose cotransporter. Re-evaluation of function and di
A;Reference number: A53582; MUID:94216314; PMID:8163506
A;Accession: A53582
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-665 <LEE>
A;Cross-references: GB:U03120; NID:g414571; PIDN:AAA19015.1; PID:g414572
C;Superfamily: proline carrier protein

  Query Match             10.3%;  Score 306;  DB 2;  Length 665;
  Best Local Similarity   23.5%;  Pred. No. 1e-14;
  Matches  155;  Conservative  105;  Mismatches  242;  Indels  158;  Gaps   29;

Qy     11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING 70
Db     32 IVIYFVVVMAVGLWAMFST-NRGTV----GGFFLAGRSMVWWPIGASLFASNIGSGHFVG 86

Qy     71 TAEAVYVPGYGLAWAQAPIGYSLS-----LILGGLFFAKPMRSK-GYVTMLDPFQQIYGK 124
Db     87 LA------GTGAAAGIAMGGFEWNALVFVVVLGWLFV--PIYIKAGVVTM----PEYLRK 134

Qy    125 RMGG------------LLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIA 172
Db    135 RFGGKRIQIYLSVLSLLLYIFTKISADIFSGAIF------INLALGLDIYLAIFILLAIT 188

Qy    173 TLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQK--PWL- 229
Db    189 ALYTITGGLAAVIYTDTLQTAIMLVGSFILTGFAFREVG----GYEAFMDKYMKAIPTLV 244

Qy    230 --GTVD-SSEVYS-WLDSFLLL---MLGGIPW----------------QAYFQRVLSS 264
Db    245 SDGNITVKEECYTPRADSFHIFRDPITGDMPWPGLIFGLSILALWYWCTDQVIVQRCLSA 304

Qy    265 SSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDP------KTTEEADM 318
Db    305 KNMSHVKAGCTLCGYLKLLPMFLMVMPGMISRILYTDKIACVLPSECKKYCGTPVGCTNI 364

Qy    319 ILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIV 378
Db    365 AYPTLVVELMPNGLRGLMLSVMMASLMSSLTSIFNSASTLFTMDIY-TKIRKGASEKELM 423

Qy    379 WVMRITVFV-FGASATAMALLTKTVYG--LWYLSSDLVYI--VIFPQLLCVLFVKGTNTY 433
Db    424 IAGRLFILVLIGISIAWVPIVQSAQSGQLFDYIQSITSYLGPPIAAVFLLAIFCKRVNEP 483

Qy    434 GAVAGYVSGLFLRI----------TG------------GEPYLYLQPLIFYPGYYPDDN 470
Db    484 GAFWGLILGFLIGISRMITEFAYGTGSCMEPSNCPKIICGVHYLYFAIILF--------- 534

Qy    471 GIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAV--VARHSEENMDK 528
Db    535 --------------AISVVTVLVISLLTKPI-------PDVHLYRLCWSLRNSTEERID- 572

Qy    529 TILVKNENIKLDELALVKPRQSMTLSSTFTNKE----------AFLDVDSSPEGSGTED 577
Db    573 --LDAGEEEPVEE----DPKDTIEIDAEAPQKEKGCFRKAYDLFCGLDQDKGPKMTKEEE 626


RESULT 7
A44432
amino acid transport protein - pig
N;Alternate names: Na+/amino acid cotransporter, SAAT1
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 20-Aug-1999
C;Accession: A44432
R;Kong, C.T.; Yet, S.F.; Lever, J.E.
J. Biol. Chem. 268, 1509-1512, 1993
A;Title: Cloning and expression of a mammalian Na+/amino acid cotransporter with se
A;Reference number: A44432; MUID:93131881; PMID:8420925
A;Accession: A44432
A;Molecule type: nucleic acid
A;Residues: 1-660 <KON>
A;Cross-references: GB:L02900; NID:g164666; PIDN:AAC37325.1; PID:g164667
A;Experimental source: kidney epithelial cell line LLC-PK1
A;Note: sequence extracted from NCBI backbone (NCBIP:122778)
C;Superfamily: proline carrier protein
C;Keywords: amino acid transport; membrane protein

  Query Match             10.2%;  Score 303.5;  DB 2;  Length 660;
  Best Local Similarity   23.2%;  Pred. No. 1.5e-14;
  Matches  141;  Conservative  103;  Mismatches  230;  Indels  135;  Gaps   26;

Qy     11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING 70
Db     32 IVIYFVVVMAVGLWAMLRT-NRGTV----GGFFLAGRDVTWWPMGASLFASNIGSGHFVG 86

Qy     71 TAEAVYVPGYGLA---WAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQIY-GKRM 126
Db     87 LAGTGAASGIAIAAFEW-----NALLLLVLGWFFVPIYIKAGVMTMPEYLRKRFGGKRL 141

Qy    127 GGLLFIPAL-------MGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVG 179
Db    142 QIYLSILSLFICVALRISSDIFSGAIF------IKLALGLDLYLAIFSLLAITAIYTITG 195

Qy    180 GLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQK--PWLGTVD---- 233
Db    196 GLASVIYTDTLQTIIMLIGSFILMGFAF----VEVGGYESFTEKYMNAIPTIVEGDNLTI 251

Qy    234 SSEVYS-WLDSFLLL---MLGGIPW----------------QAYFQRVLSSSSATYAQ 271
Db    252 SPKCYTPQGDSFHIFRDAVTGDIPWPGMIFGMTVVAAWYWCTDQVIVQRCLSGKDMSHVK 311

Qy    272 VLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKT----TEE--ADMILPIVLQ 325
Db    312 AACIMCGYLKLLPMFLMVMPGMISRILYTEKVACVVPSECVKHCGTEVGCSNYAYPLLVM 371

Qy    326 YLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITV 385
Db    372 ELMPSGLRGLMLSVMLASLMSSLTSIFNSASTLFTMDLY-TKIRKQASEKELLIAGRLFI 430

Qy    386 FVFGASATAMALLTKTVYG---LWYLSSDLVYI--VIFPQLLCVLFVKGTNTYGA----V 436
Db    431 ILLIVISIVWVPLVQVAQNGQLFHYIESISSYLGPPIAAVFLLAIFCKRVNEQGAFWGLI 490

Qy    437 AGYVSGL------FLRITG----------GEPYLYLQPLIFYPGYYPDDNGIYNQKF 477
Db    491 IGFVMGLIRMIAEFVYGTGSCLAASNCPQIICGVHYLYFALILFF--------------- 535

Qy    478 PFKTLAMVTSFLTNICISYLAK-----YLFE-----SGTLPPKLDVFDAVVARH------ 521
```

```
                 | |  : || | |     :|:       :   ::|: ||   ||
Db    536 --------VSILVVLAISLLTKPIPDVHLYRLCWALRNSTEERIDL-DAEEKRHEEAHDG 586

Qy    522 -SEENMDKT 529
           |:| ::|
Db    587 VDEDNPEET 595


RESULT 8
E83468
probable sodium/solute symporter PA1418 [imported] - Pseudomonas aeruginosa (strain PAO1
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: E83468
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: E83468
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-463 <STO>
A;Cross-references: GB:AE004571; GB:AE004091; NID:g9947360; PIDN:AAG04807.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA1418

  Query Match             10.1%;  Score 301;  DB 2;  Length 463;
  Best Local Similarity   25.1%;  Pred. No. 1.5e-14;
  Matches  115;  Conservative   86;  Mismatches  211;  Indels   46;  Gaps   15;

Qy      9 IAIIVFYLLILLVGI----WAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGF---TMTAT 61
          :|: :| :||   |:    |   |      |: | :  :| ||::|    ||   || ||
Db      1 MALDIFVVLIYAAGMIALGWYGMR-----RAKTRDD-YLVAGRNLG---PGFYLGTMAAT 51

Qy     62 WVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQI 121
           :||    ||    || |    |    ||  | ::   || |||:      |:    ::
Db     52 VLGGASTIGTVRLGYVHGISGFWLCGAIG--LGIVGLSLFLAKPLLKLKIYTVTQVLERR 109

Qy    122 YGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGGL 181
          |         :  |:  :   |    |:|  : |:  :   :|::|    :  ||: :||:
Db    110 YNPAARHASALIMLVYALMIGATSTIAIGTVMQVLFGLPFWVSILIGGGVVVLYSTIGGM 169

Qy    182 YSVAYTDVVQLFCIFVGL-WISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSW 240
          :|:  ||:||   : ||| :: :| :::       |: |: ||    :    | : :  |
Db    170 WSLTLTDIVQFLIMTVGLVFLLMPLSINDAG----GWDALVAKLPASYF---DFTAI-GW 221

Qy    241 ---LDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGAS 297
             :  ||:   |    |  :||| :: | | |:|    |   |::  :   |||
Db    222 DTIVTYFLIYFFGIFIGQDIWQRVFTARSETVAKVAGSAAGIYCVLYGMAGALIGMAAKV 281

Qy    298 TDWNQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASS 357
                |||    | :    |:::  |  |    : |  ||:||:| : :|:||:
Db    282 L--------LPD---LENVNNAFASVVEHSLPNGIRGLVIAAALAALMSTASAGLLAAST 330

Qy    358 MFARNIY-QLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIV 416
          :::  :|   :  ||   |  ||   : |    :||:   |     :: :|:
Db    331 TVTQDLLPRLRRGRGQSDNGDVHENRIATLLLGLVVLGIALVVSDVISALTVAYNLLVGG 390

Qy    417 IFPQLLCVLFVKGTNTYGAVA----GYVSGLFLRITGG 450
          :   |:  :: |   | ||:     |::: |   |  |
Db    391 MLIPLIGAIYWKRATTAGAITSMTLGFLTVLVFMIKDG 428


RESULT 9
B83988
proline transporter opuE [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: B83988
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: B83988
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-507 <STO>
A;Cross-references: GB:AP001516; GB:BA000004; NID:g10175192; PIDN:BAB06425.1; GSPDB:G
A;Experimental source: strain C-125
C;Genetics:
A;Gene: opuE
C;Superfamily: proline carrier protein

  Query Match             10.1%;  Score 299.5;  DB 2;  Length 507;
  Best Local Similarity   26.2%;  Pred. No. 2.2e-14;
  Matches  141;  Conservative   84;  Mismatches  220;  Indels   93;  Gaps   27;

Qy      5 VEGL-IAIIVFYLL-ILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATW 62
          || | :||:: ||: :||:|: :: :  : | :       : ||::       :  ::
Db      4 VEPLAVAILIAYLVALLLIGLLSS-KKSSVGMTD-----FFIAGRNLNKWTVALSAVSSG 57

Qy     63 VGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGY------VTMLD 116
             : |     |  |    || |  ||    :   |||     | : |      :|: |
Db     58 RSAWLVLGVTGTAYATGLDAVWAVA--GYITVEVF--LFFYVARRFRAYSEQTGSITIPD 113

Qy    117 PFQQIYGKR----MGGLLFIPALMGEMFWAAAIFSAL---GATISVIIDVDMHISVIISA 169
            :  :  :     ||  ||  :|   |  | : | |   |   :  : |      : ::|
Db    114 ILETRFNDKTHILRGGSAFI--IM--FFMIAYVASQLVAGGGAFATSMGVSSSTGMWVTA 169

Qy    170 LIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIG-------FTAVHA 222
            :|   ||::|| ::|: ||||   :|| | |         | || ||          :|
Db    170 VILLAYTMLGGFHAVSKTDVVQAGFMFVSLVIL-------PVVAIIGLGGFDPLLQVMHT 222

Qy    223 KYQKPWLGTVDSSEVYSWLDSFLLLMLG-GIPWQAY-FQRVLSSSSATYAQVLSFLAAFG 280
          :      |   |   : :    || :| | |   :  | :|  :    :  : :::
Db    223 EG-----GGFTSPFAFGFGAVIGLLGIGFGSPGNPHILVRYMSLKNVKEMRQAALISSVW 277

Qy    281 CLVMAIPAILIGAIGASTDWNQTAYGLPDPKTTEEAD---MILPIVLQYLCPVYISFFGL 337
            ::|   |::||  |         ||  ||       |   : | :  :  | |::  |  |
Db    278 NVLMGWGAVMIGLAG-------RAY-FPDVSLLPNGDQEQVFLMLGSEILHPLFFGFL-L 328

Qy    338 GAVSAAVMSSADSSILSASSMFARNIYQLSFRQN--ASDKEIVWVMRITVFVFGASATAM 395
          || ||:|||||| :|  || | |:|||  ||:|   | |::| : |:|  ||    :  :
Db    329 VAVLAAIMSSADSQLLVGSSAFVRDIYQKMFRRNRKLSQKKLVRLSRLTTVVFMGLSLIL 388

Qy    396 ALLTKTVYGLWYLSSDLVYIVIF----------PQLLCVLFVKGTNTYGAVAGYVSGLFL 445
          |  |   :  |         :|:|          | ||   : ||    | : | ::||
Db    389 A-FTAQEFVFW--------MVLFAFGGLGACFGPALLLSFYWKGVTRQGVLWGMIAGLLT 439

Qy    446 RITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLT----NICISYLAK 499
          |      :  ||   |  : ||   : |  |   |   |  |:      : ||   |
Db    440 VI-----LVKQQPQWTY-AFLPDVKELLNTYFFGITYEAVPGFIVATTITVVISLFTK 491


RESULT 10
A42251
nucleoside transport protein - rabbit
N;Alternate names: Na+/nucleoside cotransporter, SNST1
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 20-Aug-1999
C;Accession: A42251
R;Pajor, A.M.; Wright, E.M.
J. Biol. Chem. 267, 3557-3560, 1992
A;Title: Cloning and functional expression of a mammalian Na+/nucleoside cotranspor
A;Reference number: A42251; MUID:92156077; PMID:1740408
A;Accession: A42251
A;Molecule type: mRNA
A;Residues: 1-672 <PAJ>
A;Cross-references: GB:M84020; NID:g165550; PIDN:AAA31421.1; PID:g165551
A;Note: sequence extracted from NCBI backbone (NCBIN:82253, NCBIP:82256)
C;Superfamily: proline carrier protein
C;Keywords: membrane protein; nucleoside transport
```

Query Match 10.0%; Score 298; DB 2; Length 672;
Best Local Similarity 25.0%; Pred. No. 3.9e-14;
Matches 153; Conservative 89; Mismatches 232; Indels 138; Gaps 25;

```
Qy       9 IAII-VFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY 67
           ||:|  ::||:: ||:|:  || | |:         : || :     | :: |: :| |:
Db      26 IAVIAAYFLLVIGVGLWSMCRT-NRGTV----GGYFLAGRSMVWWPVGASLFASNIGSGH 80

Qy      68 INGTAEAVYVPGYGLAWAQAPIGYSLS----LILGGLFFAKPMRSKGYVTMLDPFQQIYG 123
             | |      | | |   |  |:  :    ::| |  ||    : | :||     |
Db      81 FVGLA------GTGAANGLAVAGFEWNALFVVLLLGWLFAPVYLTAGVITM----PQYLR 130

Qy     124 KRMGG-----------LLFIPALMGEMFWAAAIF--SALGATISVIIDVDMHISVIISA 169
           || ||             |:|   :    :: |:|  |||  |        : |||
Db     131 KRFGGHRIRLYLSVLSLFLYIFTKISVDMFSGAVFIQQALGWNI--------YASVIALL 182

Qy     170 LIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY----- 224
           |  :||: ||| :: ||| || | |  | :|   :|         |:: :  ||
Db     183 GITMVYTVTGGLAALMYTDTVQTFVIIAGAFILTGYAFHEVG----GYSGLFDKYMGAMT 238

Qy     225 -----QKPWLGTVDSSEVYSWLDSFLLL---MLGGIPW-----------------QAYF 258
             : | :| : ||      ||: ||   : | :||                     |
Db     239 SLTVSEDPAVGNISSSCYRPRPDSYHLLRDPVTGDLPWPALLLGLTIVSGWYWCSDQVIV 298

Qy     259 QRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKT----TE 314
           || |:  : |: :    |  :  |      :: | |      :: |   |:        ||
Db     299 QRCLAGRNLTHIKAGCILCGYLKLTPMFLMVMPGMISRILYPDEVACVAPEVCKRVCGTE 358

Qy     315 E--ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNA 372
           :::  | ::  | |  :    |  : ||:||| |   |:|:|:  :|| |  |  |
Db     359 VGCSNIAYPRLVVKLMPNGLRGLMLAVMLAALMSSLASIFNSSSTLFTMDIYTL--RPRA 416

Qy     373 SDKEIVWVMRITVFVFGASATAMALLTKTVYG---LWYLSSDLVYIV--IFPQLLCVLFV 427
           : |:: | |: |    | : |   : :      |: |   |:   :    : |||
Db     417 GEGELLLVGRLWVVFIVAVSVAWLPVVQAAQGGQLFDYIQSVSSYLAPPVSAVFVVALFV 476

Qy     428 KGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMV-- 485
           |  ||  | : || : :    |                    :| | | : |
Db     477 PRVNEKGAFWGLIGGLLMGLARLIP----------------------EFSFGTGSCVRP 513

Qy     486 ---TSFLTNICISYLAKYLFE-SG---------TLP-PKLDVFDAVVA-RHSEENMDKTI 530
              :||  :   | |  ||  ||         | | |:  :   | : |||:|
Db     514 SACPAFLCRVHYLYFAIVLFFCSGLLIIIVSLCTAPIPRKHLHRLVFSLRHSKE------ 567

Qy     531 LVKNENIKLDEL 542
             : |::  |||
Db     568 --EREDLDADEL 577
```

RESULT 11
S59637
glucose transport protein SGLT1, intestinal - sheep
N;Alternate names: Na+/glucose cotransporter SGLT1
C;Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C;Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 20-Aug-1999
C;Accession: S59637; S48858
R;Tarpey, P.S.; Wood, I.S.; Shirazi-Beechey, S.P.; Beechey, R.B.
Biochem. J. 312, 293-300, 1995
A;Title: Amino acid sequence and the cellular location of the Na(+)-dependent D-glucose
A;Reference number: S59637; MUID:96077158; PMID:7492327
A;Accession: S59637
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-664 <TAR>
A;Cross-references: EMBL:X82411; NID:g861072; PIDN:CAA57809.1; PID:g861073
A;Experimental source: tissue type jejunal mucosa
R;Wood, I.
submitted to the EMBL Data Library, October 1994
A;Reference number: S48858
A;Accession: S48858
A;Molecule type: mRNA
A;Residues: 1-233,'R',235-432,'V',434-466,'MR',469-664 <WOO>
A;Cross-references: EMBL:X82411
C;Superfamily: proline carrier protein

Query Match 9.9%; Score 294; DB 2; Length 664;
Best Local Similarity 23.9%; Pred. No. 7.6e-14;
Matches 127; Conservative 93; Mismatches 202; Indels 110; Gaps 23;

```
Qy      11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY--- 67
           |:::::::: ||:|| : | | |:         : || :     | :: |: :| |:
Db      32 IVIYFVVVMAVGLWAMFST-NRGTV----GGFFLAGRSMVWWPIGASLFASNIGSGHFVG 86

Qy      68 INGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSK-GYVTMLDPFQQIYGKRM 126
           : || | :   |  |        | ::|| :|   |:  | | |||     :   ||
Db      87 LAGTGAAAGIATGGFEWN----ALILVVLLGWVFV--PIYIKAGVVTM----PEYLRKRF 136

Qy     127 GG-----------LLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATL 174
           ||            :|:|   :    :: |||     |:: : :|::::: |   | |
Db     137 GGQRIQVYLSVLSLVLYIFTKISADIFSGAIF------INLALGLDLYLAIFILLAITAL 190

Qy     175 YTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDS 234
           ||: ||| :| ||| :|   : :| :|   ||          |::|   ||    : || |
Db     191 YTITGGLAAVIYTDTLQTVIMLLGSFILTGFAFHEVG----GYSAFVTKYMNA-IPTVTS 245

Qy     235 -------SEVYS-WLDSFLLL---MLGGIPW-----------------QAYFQRVLSSS 265
                 | |:   ||| :    : | :||                   |   || ||:
Db     246 YGNTTVKKECYTPRADSFHIFRDPLKGDLPWPGLIFGLTIISLWYWCTDQVIVQRCLSAK 305

Qy     266 SATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKTTEE--------AD 317
           : :: :    :  :  |:     :: | |       : |  : |   |:        :
Db     306 NMSHVKAGCIMCGYMKLLPMFLMVMPGMISRILFTEKVACTV--PSECEKYCGTKVGCTN 363

Qy     318 MILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEI 377
           :  | ::  | |  :    |  : |::|||  |   |||::|  :||    |: ||:||:
Db     364 IAYPTLVVELMPNGLRGLMLSVMLASLMSSLTSIFNSASTLFTMDIY-TKIRKKASEKEL 422

Qy     378 VWVMRITVFV-FGASATAMALLTKTVYG--LWYLSSDLVYI--VIFPQLLCVLFVKGTNT 432
           :   |: : |  | |   : ::     |    |: |   |:   |    |  :| |  |
Db     423 MIAGRLFMLVLIGVSIAWVPIVQSAQSGQLFDYIQSITSYLGPPIAAVFLLAIFCKRVNE 482

Qy     433 YGAVAGYVSGLFLRI----------TG------------GEPYLYLQPLIF 461
           ||  | : |  : :           ||             |  |||   ::|
Db     483 PGAFWGLIIGFLIGVSRMITEFAYGTGSCMEPSNCPTIICGVHYLYFAIILF 534
```

RESULT 12
A56765
sodium-glucose cotransporter homolog - human
C;Species: Homo sapiens (man)
C;Date: 08-Sep-1995 #sequence_revision 08-Sep-1995 #text_change 20-Aug-1999
C;Accession: A56765; I51890
R;Wells, R.G.; Pajor, A.M.; Kanai, Y.; Turk, E.; Wright, E.M.; Hediger, M.A.
Am. J. Physiol. 263, F459-F465, 1992
A;Title: Cloning of a human kidney cDNA with similarity to the sodium-glucose cotran
A;Reference number: A56765; MUID:93035768; PMID:1415574
A;Accession: A56765
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-672 <WEL>
A;Cross-references: GB:M95549; NID:g338052; PIDN:AAA36608.1; PID:g338053
A;Experimental source: kidney cortex
C;Superfamily: proline carrier protein
C;Keywords: transmembrane protein

Query Match 9.8%; Score 292; DB 2; Length 672;
Best Local Similarity 24.1%; Pred. No. 1.1e-13;
Matches 147; Conservative 91; Mismatches 237; Indels 136; Gaps 22;

```
Qy       8 LIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY 67
           :: |  ::||:: ||:|:  || | |:         : || :     | :: |: :| |:
Db      26 ILVIAAYFLLVIGVGLWSMCRT-NRGTV----GGYFLAGRSMVWWPVGASLFASNIGSGH 80

Qy      68 INGTAEAVYVPGYGLAWAQAPIGYSLS----LILGGLFFAKPMRSKGYVTMLDPFQQIYG 123
             | |      | | |   |  |:  :    ::| |  ||    : | :||     |
Db      81 FVGLA------GTGAASGLAVAGFEWNALFVVLLLGWLFAPVYLTAGVITM----PQYLR 130
```

```
Qy    124 KRMGG-----------LLFIPALMGEMFWAAAIF--SALGATISVIIDVDMHISVIISA 169
Db    131 KRFGGRRIRLYLSVLSLFLYIFTKISVDMFSGAVFIQQALGWNI--------YASVIALL 182

Qy    170 LIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY----- 224
Db    183 GITMIYTVTGGLAALMYTDTVQTFVILGGACILMGYAFHEVG----GYSGLFDKYLGAAT 238

Qy    225 -----QKPWLGTVDSSEVYSWLDSFLLL---MLGGIPW-----------------QAYF 258
Db    239 SLTVSEDPAVGNISSFCYRPRPDSYHLLRHPVTGDLPWPALLLGLTIVSGWYWCSDQVIV 298

Qy    259 QRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKT----TE 314
Db    299 QRCLAGKSLTHIKAGCILCGYLKLTPMFLMVMPGMISRILYPDEVACVVPEVCRRVCGTE 358

Qy    315 E--ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNA 372
Db    359 VGCSNIAYPRLVVKLMPNGLRGLMLAVMLAALMSSLASIFNSSSTLFTMDIY-TRLRPRA 417

Qy    373 SDKEIVWVMRI-TVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQLLCV----LFV 427
Db    418 GDRELLLVGRLWVVFIVVVSVAWLPVVQAAQGGQLFDYIQAVSSYLAPPVSAVFVLALFV 477

Qy    428 KGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMV-- 485
Db    478 PRVNEQGAFWGLIGGLLMGLARLIP----------------------EFSFGSGSCVQP 514

Qy    486 ---TSFLTNICISYLAKYLFE-SGTLPPKLDVFDAVV-----------ARHSEENMDKTI 530
Db    515 SACPAFLCGVHYLYFAIVLFFCSGLLTLTVSLCTAPIPRKHLHRLVFSLRHSKE------ 568

Qy    531 LVKNENIKLDE 541
Db    569 --EREDLDADE 577
```

RESULT 13
S59638
glucose transport protein SGLT1, parotid gland - sheep
N;Alternate names: Na+/glucose cotransporter SGLT1
C;Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 07-May-1999
C;Accession: S59638; S48857
R;Tarpey, P.S.; Wood, I.S.; Shirazi-Beechey, S.P.; Beechey, R.B.
Biochem. J. 312, 293-300, 1995
A;Title: Amino acid sequence and the cellular location of the Na(+)-dependent D-glucose
A;Reference number: S59637; MUID:96077158; PMID:7492327
A;Accession: S59638
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 1-664 <TAR>
A;Cross-references: EMBL:X82410
A;Experimental source: clone SGLTB; tissue type parotid gland
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1994
C;Superfamily: proline carrier protein
C;Keywords: transmembrane protein

Query Match 9.7%; Score 288; DB 2; Length 664;
Best Local Similarity 23.7%; Pred. No. 2.1e-13;
Matches 126; Conservative 93; Mismatches 203; Indels 110; Gaps 23;

```
Qy     11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY--- 67
Db     32 IVIYFVVVMAVGLWHMFST-NRGTV----GGFFLAGRSMVWWPIGASLFASNIGSGHFVG 86

Qy     68 INGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSK-GYVTMLDPFQQIYGKRM 126
Db     87 LAGTGAAAGIATGGFEWN----ALILVVLLGWVFV--PIYIKAGVVTM----PEYLRKRF 136

Qy    127 GG-----------LLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATL 174
Db    137 GGQRIQVYLSVLSLVLYIFTKISADIFSGAIF------INLALGLDLYLAIFILLAITAL 190

Qy    175 YTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDS 234
Db    191 YTITGGLAAVIYTDTLQTVIMLLGSFILTGFAFHEVG----GYSAFVTKYMNA-IPTVTS 245

Qy    235 -------SEVYS-WLDSFLLL---MLGGIPW-----------------QAYFQRVLSSS 265
Db    246 YGNTTVKKECYTPRADSFHIFRDPLKGDLPWPGLIFGLTIISLWYWCTDQVIVQRCLSAK 305

Qy    266 SATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKTTEE--------AD 317
Db    306 NMSHVKAGCIMCGYMKLLPMFLMVMPGMISRILFTEKVACTV--PSECEKYCGTKVGCTN 363

Qy    318 MILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEI 377
Db    364 IAYPTLVVELMPNGLRGLMLSVMLASLMSSLTSIFNSASTLFTMDIY-TKIRKKASEKEL 422

Qy    378 VWVMRITVFV-FGASATAMALLTKTVYG--LWYLSSDLVYI--VIFPQLLCVLFVKGTNT 432
Db    423 MIAGRLFMLVLIGVSIAWVPIVQSAQSGQLFDYIQSITSYLGPPIAAVFLLAIFCKRVNE 482

Qy    433 YGAVAGYVSGLFLRI----------TG------------GEPYLYLQPLIF 461
Db    483 PGAFWGLIIGFLIGVSRMITEFAYGTGSCMEPSNCPTIICGVHYLYFAIILF 534
```

RESULT 14
H71097
hypothetical protein PH1044 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C;Accession: H71097
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; S
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Og
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophil
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: H71097
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-491 <KAW>
A;Cross-references: GB:AP000004; NID:g3236131; PIDN:BAA30142.1; PID:g3257459
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by Genb
C;Genetics:
A;Gene: PH1044
C;Superfamily: proline carrier protein

Query Match 9.6%; Score 286; DB 2; Length 491;
Best Local Similarity 22.9%; Pred. No. 2.1e-13;
Matches 125; Conservative 83; Mismatches 197; Indels 142; Gaps 20;

```
Qy      8 LIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY 67
Db     17 LIIVLGWVFLSLIVGVMAGIKRKFT------LEGYLVSGRTLGLIFLYVLMAGEIYSAYA 70

Qy     68 INGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFA----KPMRSKGYVTMLDPFQQIYG 123
Db     71 FLGTGGWAYSYGMPIMYA---IGYGALAYSFGYFYARYVWKAGKAFGCVTQADYFQVRYN 127

Qy    124 KRMGGLLFIPALMGEMF---WAAAIFSALGATISV--IIDVDMHISVIISALIATLYTLV 178
Db    128 SK--ALAVLVALIGIIFNIPYLQLQLQGLGYIVHVGSLGSITPKAGIVIGMIIMMIYVYT 185

Qy    179 GGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVY 238
Db    186 SGLRGISWTNLLQATLMFIVAWV-VLFTIPFKQFGGIGELFKTLAQTKP---------- 233

Qy    239 SWLDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLV--MAIPAILIGAIGA 296
Db    234 ---DHLILHPPLGISW----------YVSTL-ILSGLGFFMYPQLYPSI------- 268
```

```
Qy    297 STDWNQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVS---------- 341
          ||  | ||  :   ::||:   :: || :: | : |:
Db    269 --------YGARDLKTLKRNYVLLPLYSIFMIPVILAGFTVAALGIKLSAPDEAVLKAVE 320

Qy    342 ----------------AAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITV 385
                         ||  |:| : :|| : : ::|:| :: :  |||||:| | ||:|
Db    321 ITYPSWVLGVVGAAGFAAAASTASAILLSLAGLLSKNLYAIA-KPTASDKELVLVSRISV 379

Qy    386 FVFGASATAMALLTKTVYGLWYLSSDLVYIVI---------FPQLLCVLFVKGTNTY--- 433
          : |  |  :||         |    || :::          ||  :  || |  | |
Db    380 ILLGLLAMGLAL---------YSPGRLVSLLLLAYAGMTQMFPGAVFGLFWKRMNKYATG 430

Qy    434 -GAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNI 492
          | :|| ::  :||:       |  ||          ||:     |    :: : :  :
Db    431 TGIIAGLITVAYLRLV-----LKKNPL----------GIH-----FGLWGLLVNIIVTL 469

Qy    493 CISYLAK 499
          ::|| |
Db    470 IVAYLTK 476
```

RESULT 15
H69670
sodium/proline symporter opuE - Bacillus subtilis
N;Alternate names: proline transporter opuE
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C;Accession: H69670; T44450
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: H69670
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-492 <KUN>
A;Cross-references: GB:Z99107; GB:AL009126; NID:g2632866; PIDN:CAB12486.1; PID:g2632980
A;Experimental source: strain 168
A;Borriss, R.
submitted to the EMBL Data Library, June 1997
A;Reference number: Z22776
A;Accession: T44450
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-492 <BOR>
A;Cross-references: EMBL:AF011545; PIDN:AAB72182.1
C;Genetics:
A;Gene: opuE
A;Map position: 56 degree
C;Function:
A;Description: catalyzes the uptake of proline by a Na+-dependent transport mechanism
C;Superfamily: proline carrier protein
C;Keywords: proline transport; sodium transport; symport system; transmembrane protein
F;44-71/Domain: transmembrane #status predicted <TM1>
F;126-145/Domain: transmembrane #status predicted <TM2>
F;161-183/Domain: transmembrane #status predicted <TM3>
F;189-208/Domain: transmembrane #status predicted <TM4>
F;231-253/Domain: transmembrane #status predicted <TM5>
F;272-295/Domain: transmembrane #status predicted <TM6>
F;311-347/Domain: transmembrane #status predicted <TM7>
F;366-386/Domain: transmembrane #status predicted <TM8>
F;391-417/Domain: transmembrane #status predicted <TM9>
F;422-438/Domain: transmembrane #status predicted <TM10>
F;452-470/Domain: transmembrane #status predicted <TM11>

Query Match 9.6%; Score 285; DB 2; Length 492;
Best Local Similarity 22.1%; Pred. No. 2.5e-13;
Matches 118; Conservative 97; Mismatches 214; Indels 106; Gaps 18;

```
Qy      5 VEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVG 64
          :| :|:: :::: :||:| :|  :| :              ::||| :|  |   :  | :
Db      3 IEIIISLGIYFIAMLLIGWYAFKKTTDIND-------YMLGGRGLGPFVTALSAGAADMS 55

Qy     65 GGYINGTAEAVYVPGYGLAWAQAPI---GYSLSLILGGLFFAKPMRSKGYVTMLDPFQQI 121
          |  : |   |::  |    |     :   ||  |:|     |    :   :|: | | :
Db     56 GWMLMGVPGAMFATGLSTLWLALGLTIGAYSNYLLLAPRLRAYTEAADDAITIPDFFDKR 115

Qy    122 YGKRMGGLLFIPALMGEMFWAAAIFSAL---GATISVIIDVDMHISVIISALIATLYTLV 178
          :     |  : ||:  :|:     | :  |          |  : : ::  :  ||||
Db    116 FQHSSSLLKIVSALIIMIFFTLYTSSGMVSGGRLFESAFGADYKLGLFLTTAVVVLYTLF 175

Qy    179 GGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVY 238
          ||  :|: || ||   :|  | : ||          : ||  |     |    :|:  :
Db    176 GGFLAVSLTDFVQGAIMFAAL-VLVPI---------VAFT--HVGGVAPTFHEIDAVNPH 223

Qy    239 SWLDSF----LLLMLGGIPWQAYFQRVLSSSSATYAQ---VLSFLA-------------- 277
            || |    :: ::  : |   :           | |   :: |:|
Db    224 -LLDIFKGASVISIISYLAWGLGY----------YGQPHIIVRFMAIKDIKDLKPARRIG 272

Qy    278 AFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGL 337
          ::  : ::| | ||         ||         :: :||  |  : |     |: | |
Db    273 MSWMIITVLGSVLTGLIG-------VAYAHKFGVAVKDPEMIFIIFSKILFHPLITGFLL 325

Qy    338 GAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMAL 397
          |: ||:|||  | :|  :|    ::|:  ||: |||||:| : |::| |    |  ::|
Db    326 SAILAAIMSSISSQLLVTASAVTEDLYRSFFRRKASDKELVMIGRLSVLVIAVIAVLLSL 385

Qy    398 LTKTVYGLWYLSSDLVYIVIF----------PQLLCVLFVKGTNTYGAVAGYVSG---LF 444
                :| :: :| :          | :|  |: |  | :||:|  : |   :
Db    386 NP---------NSTILDLVGYAWAGFGSAFGPAILLSLYWKRMNEWGALAAMIVGAATVL 436

Qy    445 LRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAK 499
          : || |                |:| :  |      : | :  | :| : |
Db    437 IWITTG--------------LAKSTGVY-EIIP----GFILSMIAGIIVSMITK 471
```

Search completed: March 31, 2003, 18:36:24
Job time : 52 secs

OM protein - protein search, using sw model

Run on: March 31, 2003, 18:35:31 ; Search time 593 Seconds
(without alignments)
57.436 Million cell updates/sec

Title: US-10-069-541-6
Perfect score: 2972
Sequence: 1 MAFHVEGLIAIIVFYLLILL.....EAFLDVDSSPEGSGTEDNLQ 580

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 237916 seqs, 58723674 residues

Total number of hits satisfying chosen parameters: 237916

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	311.5	10.5	675	10	US-09-733-630-2	Sequence 2, Appli
2	306	10.3	664	9	US-10-119-988-12	Sequence 12, Appl
3	298.5	10.0	675	9	US-09-928-530-2	Sequence 2, Appli
4	298.5	10.0	675	9	US-10-162-012-27	Sequence 27, Appl
5	295	9.9	596	9	US-10-119-988-8	Sequence 8, Appli
6	292	9.8	672	9	US-09-928-530-5	Sequence 5, Appli
7	292	9.8	672	9	US-10-162-012-30	Sequence 30, Appl
8	277.5	9.3	596	9	US-10-237-859-2	Sequence 2, Appli
9	277.5	9.3	596	10	US-09-740-026A-2	Sequence 2, Appli
10	277.5	9.3	643	9	US-10-119-988-5	Sequence 5, Appli
11	277	9.3	596	9	US-10-237-859-4	Sequence 4, Appli
12	277	9.3	596	10	US-09-740-026A-4	Sequence 4, Appli
13	272.5	9.2	524	9	US-09-738-626-6949	Sequence 6949, Ap
14	269.5	9.1	664	9	US-10-119-988-2	Sequence 2, Appli
15	260	8.7	674	9	US-10-173-123-9	Sequence 9, Appli
16	260	8.7	681	9	US-10-173-123-7	Sequence 7, Appli
17	245	8.2	738	9	US-10-173-123-13	Sequence 13, Appl
18	245	8.2	745	9	US-10-173-123-11	Sequence 11, Appl
19	236.5	8.0	610	12	US-10-060-974-2	Sequence 2, Appli
20	235	7.9	447	9	US-09-928-530-4	Sequence 4, Appli
21	235	7.9	447	9	US-10-119-988-10	Sequence 10, Appl
22	235	7.9	447	9	US-10-162-012-29	Sequence 29, Appl
23	227	7.6	551	9	US-10-216-355-4	Sequence 4, Appli
24	219	7.4	551	9	US-09-738-626-4431	Sequence 4431, Ap
25	218.5	7.4	618	9	US-10-216-355-2	Sequence 2, Appli
26	218	7.3	618	10	US-09-995-007-2	Sequence 2, Appli
27	141.5	4.8	267	10	US-09-864-761-46821	Sequence 46821, A
28	137	4.6	1019	9	US-09-738-626-6499	Sequence 6499, Ap
29	125	4.2	476	10	US-09-815-242-10014	Sequence 10014, A
30	121.5	4.1	459	10	US-09-815-242-12844	Sequence 12844, A
31	120	4.0	786	10	US-09-815-242-10103	Sequence 10103, A
32	116	3.9	446	10	US-09-815-242-13395	Sequence 13395, A
33	116	3.9	751	10	US-09-815-242-14001	Sequence 14001, A
34	115.5	3.9	437	10	US-09-815-242-11837	Sequence 11837, A
35	115	3.9	453	9	US-09-738-626-4026	Sequence 4026, Ap
36	114.5	3.9	452	10	US-09-815-242-13983	Sequence 13983, A
37	114	3.8	697	10	US-09-881-752A-76	Sequence 76, Appl
38	113.5	3.8	482	10	US-09-815-242-5401	Sequence 5401, Ap
39	113.5	3.8	488	10	US-09-815-242-12564	Sequence 12564, A
40	113	3.8	426	9	US-09-738-626-6036	Sequence 6036, Ap
41	113	3.8	517	9	US-09-815-923-16	Sequence 16, Appl
42	112	3.8	615	9	US-09-738-626-6987	Sequence 6987, Ap
43	111	3.7	485	10	US-09-815-242-13712	Sequence 13712, A
44	110.5	3.7	549	9	US-09-738-626-3926	Sequence 3926, Ap
45	110.5	3.7	549	10	US-09-946-763-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-733-630-2
; Sequence 2, Application US/09733630
; Patent No. US20020034799A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Freidrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Abuin, Alejandro
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020034799A1el Human Transporter Protein and
; TITLE OF INVENTION: Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0106-USA
; CURRENT APPLICATION NUMBER: US/09/733,630
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: US 60/170,137
; PRIOR FILING DATE: 1999-12-10
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 675
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-733-630-2

Query Match 10.5%; Score 311.5; DB 10; Length 675;
Best Local Similarity 23.0%; Pred. No. 1.3e-18;
Matches 152; Conservative 112; Mismatches 235; Indels 161; Gaps 28;

```
Qy       2 AFHVEGL----IAIIVFY-LLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGF 56
           ||  :||    ||::| | | :| ||:|:  :||      :  :   : | |:     |
Db      18 AFPQKGLEPGDIAVLVLYFLFVLAVGLWSTVKTK-----RDTVKGYFLAGGDMVWWPVGA 72

Qy      57 TMTATWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLI-LGGLF-------FAKPMRS 108
           :: |: || |:             ||| : |  | |:|   | |||          |:
Db      73 SLFASNVGSGHF-----------IGLAGSSAATGISVSAYELNGLFSVLMLAWIFLPIYI 121

Qy     109 KGYVTMLDPFQQIYGKRMGGLLFIPALMGEMFWAAAIFSAL------GAT-ISVIIDVDM 161
             | || :  : :   || ||:  || ::  ::    ||: :      ||  |   : :|:
Db     122 AGQVTTMPEYLR---KRFGGIR-IPIILAVLYLFIYIFTKISVDMYAGAIFIQQSLHLDL 177
```

```
Qy    162 HISVIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVH 221
          :::::    |  :||: ||| :| ||| :|   : :|    : :  |  ||   |   :
Db    178 YLAIVGLLAITAVYTVAGGLAAVIYTDALQTLIMLIGALTLMGY--SFAAVG--GMEGLK 233

Qy    222 AKY------------------------------QKPWLGTVDSSEVYS-WLDSFLLLML 249
           ||                                || | :    : | |
Db    234 EKYFLALASNRSENSSCGLPREDAFHIFRDPLTSDLPWPGVLFGMSIPSLWY-------- 285

Qy    250 GGIPW---QAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYG 306
            |   |   || |:: : ::|:  : :||:  ::     :  | :      :| |
Db    286 ----WCTDQVIVQRTLAAKNLSHAKGGALMAAYLKVLPLFIMVFPGMVSRILFPDQVA-- 339

Qy    307 LPDPKTTEE--------ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSM 358
            ||:  ::        :|:  | ::  | |  :    :  : ||:|||  |   |||::
Db    340 CADPEICQKICSNPSGCSDIAYPKLVLELLPTGLRGLMMAVMVAALMSSLTSIFNSASTI 399

Qy    359 FARNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLW----------- 406
          |  :::    |  ||:||:: | |: |           ||   |  ||
Db    400 FTMDLWN-HLRPRASEKELMIVGRVFV----------LLLVLVSILWIPVVQASQGGQL 447

Qy    407 --YLSSDLVYI-----VIFPQLLCVLFVKGTNTYGAVAGYVSGLFLRITG-GEPYLYLQP 458
          |: |   |:      |:|   :   | | ||  ||  | :||| | :      ::|:||
Db    448 FIYIQSISSYLQPPVAVVF---IMGCFWKRTNEKGAFWGLISGLLLGLVRLVLDFIYVQP 504

Qy    459 LIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDV----- 513
                ||:  :  :   :   :|: | :| | :| :: :     | ||  ::
Db    505 RC----DQPDERPVLVKSIHYLYFSMILSTVTLITVSTVSWF-----TEPPSKEMVSHLT 555

Qy    514 ----FDAVVARHSEENMDKTILVKNENIKLD-------ELALVKPRQSMTLSSTFTNKEA 562
              | || :          |  ::|   :        :  :|:   | | |   | |::
Db    556 WFTRHDPVVQKEQAPPAAPLSLTLSQNGMPEASSSSSVQFEMVQENTSKTHSCDMTPKQS 615
```

RESULT 2
US-10-119-988-12
; Sequence 12, Application US/10119988
; Publication No. US20030054453A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A.J.
; APPLICANT: Chen, Hong
; APPLICANT: Millennium Pharmaceuticals Inc.
; TITLE OF INVENTION: 68723, Sodium/Glucose Cotransporter
; TITLE OF INVENTION: Family Members and Uses Therefor
; FILE REFERENCE: MPI01-103P1RNM
; CURRENT APPLICATION NUMBER: US/10/119,988
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 60/282,764
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 664
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-119-988-12

Query Match 10.3%; Score 306; DB 9; Length 664;
Best Local Similarity 22.8%; Pred. No. 3.8e-18;
Matches 148; Conservative 104; Mismatches 218; Indels 178; Gaps 30;

```
Qy     11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING 70
          |:::::::: ||:|| : | | |:         : || :     | :: |: :| |:  |
Db     32 IVIYFVVVMAVGLWAMFST-NRGTV----GGFFLAGRSMVWWPIGASLFASNIGSGHFVG 86

Qy     71 TAEAVYVPGYGLAWAQAPIGYS-----LSLILGGLFFAKPMRSK-GYVTMLDPFQQIYGK 124
          |      | | |   |  |:      | ::|| ||   |:  | | |||     :   |
Db     87 LA------GTGAASGIAIGGFEWNALVLVVVLGWLFV--PIYIKAGVVTM----PEYLRK 134

Qy    125 RMGG------------LLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIA 172
          | ||           ||:|   :    :: ||| |:: : :::::: :   |
Db    135 RFGGQRIQVYLSLLSLLLYIFTKISADIFSGAIF-----INLALGLNLYLAIFLLLAIT 188

```
Qy    173 TLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQK--PWL- 229
          |||: ||| :| ||| :|   : ||  |   ||          |: |   || |  | :
Db    189 ALYTITGGLAAVIYTDTLQTVIMLVGSLILTGFAFHEVG----GYDAFMEKYMKAIPTIV 244

Qy    230 ---GTVDSSEVYS-WLDSFLLL---MLGGIPW-----------------QAYFQRVLSS 264
              |    : |:   ||| :    : | :||                   |   || ||:
Db    245 SDGNTTFQEKCYTPRADSFHIFRDPLTGDLPWPGFIFGMSILTLWYWCTDQVIVQRCLSA 304

Qy    265 SSATYAQ----VLSFLAAFGCLVMAIPAIL--------IGAI---------GASTDWNQT 303
          : :: :    :  :|      :| :| ::        |  :          |
Db    305 KNMSHVKGGCILCGYLKLMPMFIMVMPGMISRILYTEKIACVVPSECEKYCGTKVGCTNI 364

Qy    304 AYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNI 363
          ||               | ::  | |  :    |  : |::|||  |   |||::|  :|
Db    365 AY--------------PTLVVELMPNGLRGLMLSVMLASLMSSLTSIFNSASTLFTMDI 409

Qy    364 YQLSFRQNASDKEIVWVMRITVFV-FGASATAMALLTKTVYG--LWYLSSDLVYI--VIF 418
          |    |: ||:||::   |: : |  | |   : ::     |    |: |   |:   |
Db    410 Y-AKVRKRASEKELMIAGRLFILVLIGISIAWVPIVQSAQSGQLFDYIQSITSYLGPPIA 468

Qy    419 PQLLCVLFVKGTNTYGAVAGYVSGLFLRI----------TG-------------GEPYLY 455
            | :| |  |  ||  | : || : |             ||              |  |||
Db    469 AVFLLAIFWKRVNEPGAFWGLILGLLIGISRMITEFAYGTGSCMEPSNCPTIICGVHYLY 528

Qy    456 LQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFD 515
            ::|                          ||:| : || | | :       | : ::
Db    529 FAIILF----------------------AISFITIVVISLLTKPI-------PDVHLYR 558

Qy    516 AV--VARHSEENMDKTILVKNENIKLDELALVKPRQSMTLSSTFTNKE 561
            :    || :|  :  : |||:        |:::: : :    |:
Db    559 LCWSLRNSKEERID--LDAEEENIQ------EGPKETIEIETQVPEKK 598
```

RESULT 3
US-09-928-530-2
; Sequence 2, Application US/09928530
; Patent No. US20020156002A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A. J.
; APPLICANT: Silos-Santiago, Inmaculada
; TITLE OF INVENTION: 32620, A NOVEL HUMAN SODIUM-SUGAR
; TITLE OF INVENTION: SYMPORTER FAMILY MEMBER AND USES THEREOF
; FILE REFERENCE: 10446-080001
; CURRENT APPLICATION NUMBER: US/09/928,530
; CURRENT FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/227,068
; PRIOR FILING DATE: 2000-08-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 675
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-928-530-2

Query Match 10.0%; Score 298.5; DB 9; Length 675;
Best Local Similarity 22.7%; Pred. No. 1.7e-17;
Matches 149; Conservative 115; Mismatches 238; Indels 155; Gaps 29;

```
Qy      2 AFHVEGL----IAIIVFY-LLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGF 56
          ||  :||    ||::| | | :| ||:|:  :||     :     |  :   ||
Db     18 AFPQKGLEPGDIAVLVLYFLFVLAVGLWSTVKTKR----DTVKGYFLAEGNMVWWPVGA- 72

Qy     57 TMTATWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLS-----LILGGLFFAKPMRSKGY 111
          :: |: || |:  | |       | | |    :   | |:     |:|  :|  :|   |
Db     73 SLFASNVGSGHFIGLA------GSGAATGISVSAYELNGLFSVLMLAWIFL--PIYIAGQ 124

Qy    112 VTMLDPFQQIYGKRMGGLLFIPALMGEMFWAAAIFSALGATI---SVIID----VDMHIS 164
          || :  : :   || ||:  || ::  ::    ||: :   :   :: |     :|::::
Db    125 VTTMPEYLR---KRFGGIR-IPIILAVLYLFIYIFTKISVDMYAGAIFIQQSSHLDLYLA 180

Qy    165 VIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY 224
          ::    |  :||: ||| :| ||| :|   : :|    : :  |  ||   |   :  ||
```
```

```
Db    181 IVGLLAITAVYTVAGGLAAVIYTDALQTLIMLIGALTLMGY--SFAAVG--GMEGLKEKY 236

Qy    225 ------------------------------QKPWLGTVDSSEVYS-WLDSFLLLMLGGI 252
                                           || | :    : | |
Db    237 FLALASNRSENSSCGLPREDAFHIFRDPLTSDLPWPGVLFGMSIPSLWY----------- 285

Qy    253 PW---QAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPD 309
           |   |   || |:: : ::|:  : :||:  ::     :  | :      :| |    |
Db    286 -WCTDQVIVQRTLAAKNLSHAKGGALMAAYLKVLPLFIMVFPGMVSRILFPDQVA--CAD 342

Qy    310 PKTTEE--------ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFAR 361
          |:  ::        :|:  | ::  | |  :    :  : ||:|||  |   |||::|
Db    343 PEICQKICSNPSGCSDIAYPKLVLELLPTGLRGLMMAVMVAALMSSLTSIFNSASTIFTM 402

Qy    362 NIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLW--------------Y 407
          :::    |  ||:||:: | |: |            ||   |  ||              |
Db    403 DLWN-HLRPRASEKELMIVGRVFV-----------LLLVLVSILWIPVVQASQGGQLFIY 450

Qy    408 LSSDLVYI-----VIFPQLLCVLFVKGTNTYGAVAGYVSGLFLRITG-GEPYLYLQPLIF 461
          : |   |:     |:|   :   | | ||  ||  | :||| | :      ::|:||
Db    451 IQSISSYLQPPVAVVF---IMGCFWKRTNEKGAFWGLISGLLLGLVRLVLDFIYVQPRC- 506

Qy    462 YPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDV-------- 513
              ||:  :  :   :   :|: | :| | :| :: :     | ||  ::
Db    507 ---DQPDERPVLVKSIHYLYFSMILSTVTLITVSTVSWF-----TEPPSKEMVSHLTWFT 558

Qy    514 -FDAVVARHSEENMDKTILVKNENIKLD-------ELALVKPRQSMTLSSTFTNKEA 562
           | || :          |  ::|   :       :  :|:   | | |   | |::
Db    559 RHDPVVQKEQAPPAAPLSLTLSQNGMPEASSSSSVQFEMVQENTSKTHSCDMTPKQS 615
```

RESULT 4
US-10-162-012-27
; Sequence 27, Application US/10162012
; Publication No. US20030051660A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A.J.
; APPLICANT: Silos-Santiago, Inmaculada
; APPLICANT: Gu, Wei
; TITLE OF INVENTION: NOVEL HUMAN ION CHANNEL AND TRANSPORTER FAMILY MEMBERS
; FILE REFERENCE: 10448-190001
; CURRENT APPLICATION NUMBER: US/10/162,012
; CURRENT FILING DATE: 2002-06-04
; PRIOR APPLICATION NUMBER: US 60/209,845
; PRIOR FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: US 09/875,321
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: PCT/US01/18340
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/209,257
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 09/875,423
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: PCT/US01/18398
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/209,238
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 09/875,363
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: PCT/US01/18247
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/227,068
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: US 09/928,530
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/25475
; PRIOR FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: US 60/226,770
; PRIOR FILING DATE: 2000-08-21
; PRIOR APPLICATION NUMBER: US 09/934,421
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: PCT/US01/26096
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/279,281
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 10/109,029
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: PCT/US02/09728
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 60/290,288
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US (not assigned)
; PRIOR FILING DATE: 2002-05-13
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 675
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-162-012-27

Query Match 10.0%; Score 298.5; DB 9; Length 675;
Best Local Similarity 22.7%; Pred. No. 1.7e-17;
Matches 149; Conservative 115; Mismatches 238; Indels 155; Gaps 29;

```
Qy      2 AFHVEGL----IAIIVFY-LLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGF 56
            ||  :||    ||::| | | :| ||:|:  :||     :      :  |  :   ||
Db     18 AFPQKGLEPGDIAVLVLYFLFVLAVGLWSTVKTKR----DTVKGYFLAEGNMVWWPVGA- 72

Qy     57 TMTATWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLS-----LILGGLFFAKPMRSKGY 111
           :: |: || |:  | |      | | |   :   | |:     |:|  :|   |:   |
Db     73 SLFASNVGSGHFIGLA------GSGAATGISVSAYELNGLFSVLMLAWIFL--PIYIAGQ 124

Qy    112 VTMLDPFQQIYGKRMGGLLFIPALMGEMFWAAAIFSALGATI---SVIID----VDMHIS 164
           || :  : :   || ||:  || ::  ::    ||: :   :   :: |     :|::::
Db    125 VTTMPEYLR---KRFGGIR-IPIILAVLYLFIYIFTKISVDMYAGAIFIQQSSHLDLYLA 180

Qy    165 VIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY 224
           ::    |  :||: ||| :| ||| :|   : :|    : :  |  ||   |   :  ||
Db    181 IVGLLAITAVYTVAGGLAAVIYTDALQTLIMLIGALTLMGY--SFAAVG--GMEGLKEKY 236

Qy    225 ------------------------------QKPWLGTVDSSEVYS-WLDSFLLLMLGGI 252
                                           || | :    : | |
Db    237 FLALASNRSENSSCGLPREDAFHIFRDPLTSDLPWPGVLFGMSIPSLWY----------- 285

Qy    253 PW---QAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPD 309
           |   |   || |:: : ::|:  : :||:  ::     :  | :      :| |    |
Db    286 -WCTDQVIVQRTLAAKNLSHAKGGALMAAYLKVLPLFIMVFPGMVSRILFPDQVA--CAD 342

Qy    310 PKTTEE--------ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFAR 361
          |:  ::        :|:  | ::  | |  :    :  : ||:|||  |   |||::|
Db    343 PEICQKICSNPSGCSDIAYPKLVLELLPTGLRGLMMAVMVAALMSSLTSIFNSASTIFTM 402

Qy    362 NIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLW--------------Y 407
          :::    |  ||:||:: | |: |            ||   |  ||              |
Db    403 DLWN-HLRPRASEKELMIVGRVFV-----------LLLVLVSILWIPVVQASQGGQLFIY 450

Qy    408 LSSDLVYI-----VIFPQLLCVLFVKGTNTYGAVAGYVSGLFLRITG-GEPYLYLQPLIF 461
          : |   |:     |:|   :   | | ||  ||  | :||| | :      ::|:||
Db    451 IQSISSYLQPPVAVVF---IMGCFWKRTNEKGAFWGLISGLLLGLVRLVLDFIYVQPRC- 506

Qy    462 YPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDV-------- 513
              ||:  :  :   :   :|: | :| | :| :: :     | ||  ::
Db    507 ---DQPDERPVLVKSIHYLYFSMILSTVTLITVSTVSWF-----TEPPSKEMVSHLTWFT 558

Qy    514 -FDAVVARHSEENMDKTILVKNENIKLD-------ELALVKPRQSMTLSSTFTNKEA 562
           | || :          |  ::|   :       :  :|:   | | |   | |::
Db    559 RHDPVVQKEQAPPAAPLSLTLSQNGMPEASSSSSVQFEMVQENTSKTHSCDMTPKQS 615
```

RESULT 5
US-10-119-988-8
; Sequence 8, Application US/10119988
; Publication No. US20030054453A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A.J.

```
;  APPLICANT:  Chen, Hong
;  APPLICANT:  Millennium Pharmaceuticals Inc.
;  TITLE OF INVENTION: 68723, Sodium/Glucose Cotransporter
;  TITLE OF INVENTION:  Family Members and Uses Therefor
;  FILE REFERENCE: MPI01-103P1RNM
;  CURRENT APPLICATION NUMBER: US/10/119,988
;  CURRENT FILING DATE:  2002-04-10
;  PRIOR APPLICATION NUMBER: 60/282,764
;  PRIOR FILING DATE: 2001-04-10
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-10-119-988-8

  Query Match             9.9%;  Score 295;  DB 9;  Length 596;
  Best Local Similarity   24.2%;  Pred. No. 2.9e-17;
  Matches  131;  Conservative   91;  Mismatches  178;  Indels  142;  Gaps   23;

Qy       4 HVEG-------LIAIIVFYLLILLVGIWAAWR-TKNSGSAEERSEAIIVGGRDIGLLVGG 55
Db      10 HVPGSQLSVTDIIVISVYFALNVAVGIWSACRANKNTVS------GYFLAGRDMAWWPIG 63

Qy      56 FTMTATWVGGGYI-----NGTAEAVYVPGY---------GLAWAQAPIGYSLSLILGGLF 101
Db      64 ASLFASSEGSGLFVGLAGSGAAGGLAVAGFEWNATYVLLALAWVFVPIYISSEI------ 117

Qy     102 FAKPMRSKGYVTMLDPFQQIY-GKRMGGLLFIPALMGEMFWAAAIFSALGAT-ISVIIDV 159
Db     118 ----------VTLPEYIQKRFGGQRIRTYLSVLSLMLSVFTKISIDLYAGALFVHICLGW 167

Qy     160 DMHISVIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWI------------------ 201
Db     168 NFYLSTILTLAITALYTIAGGLATVIYTDALQTIIMVVGAVILAVKAFNQIGGYEQLEAA 227

Qy     202 --------SVPFALSHPAVADIGFTAVH-----AKYQKPWLGTVDSSEVYSWLDSFLLLM 248
Db     228 YAQAIPSRTIPNTTCHLPRAD----AMHMFRDPSTGDLPWTGM-----------TFGLTI 272

Qy     249 LGGIPW---QAYFQRVLSSSSATYAQVLSFLAAF------GCLVM--------------- 284
Db     273 MATWYWCTDQVIVQRSLSARNLNHAKAGSILASYLKMLPMGLMIMPGMISRVLFPDDVGC 332

Qy     285 AIPAILIGAIGASTDWNQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAV 344
Db     333 VVPSECLRACGAEIGCSNIAY----PKLVME---LMPIGLRGLM--------IAVMMAAL 377

Qy     345 MSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITVFV-FGASATAMALLTKTVY 403
Db     378 LSSLTSIFNSSSTLFTMDIWR-QLRPSAGERELLLVGRLVIVVLIGVSVAWIPVLQGSNS 436

Qy     404 GLWYLSSDLVYIVIFPQLLCV----LFVKGTNTYGAVAGYVSGLF---LRITGGEPYLYL 456
Db     437 GQLFIYMQSVTSSLAPPVTAIFILGIFWRRANEQGAFWGLMAGLVVGALRLV--LEFLYP 494

Qy     457 QP 458
Db     495 EP 496


RESULT 6
US-09-928-530-5
; Sequence 5, Application US/09928530
; Patent No. US20020156002A1
; GENERAL INFORMATION:
;  APPLICANT: Curtis, Rory A. J.
;  APPLICANT:  Silos-Santiago, Inmaculada
;  TITLE OF INVENTION: 32620, A NOVEL HUMAN SODIUM-SUGAR
;  TITLE OF INVENTION:  SYMPORTER FAMILY MEMBER AND USES THEREOF
;  FILE REFERENCE: 10446-080001
;  CURRENT APPLICATION NUMBER: US/09/928,530
;  CURRENT FILING DATE:  2001-08-13
;  PRIOR APPLICATION NUMBER: 60/227,068
;  PRIOR FILING DATE: 2000-08-22
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 672
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-928-530-5

  Query Match             9.8%;  Score 292;  DB 9;  Length 672;
  Best Local Similarity   24.1%;  Pred. No. 6.3e-17;
  Matches  147;  Conservative   91;  Mismatches  237;  Indels  136;  Gaps   22;

Qy       8 LIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY 67
Db      26 ILVIAAYFLLVIGVGLWSMCRT-NRGTV----GGYFLAGRSMVWWPVGASLFASNIGSGH 80

Qy      68 INGTAEAVYVPGYGLAWAQAPIGYSLS----LILGGLFFAKPMRSKGYVTMLDPFQQIYG 123
Db      81 FVGLA------GTGAASGLAVAGFEWNALFVVLLLGWLFAPVYLTAGVITM----PQYLR 130

Qy     124 KRMGG------------LLFIPALMGEMFWAAAIF--SALGATISVIIDVDMHISVIISA 169
Db     131 KRFGGRRIRLYLSVLSLFLYIFTKISVDMFSGAVFIQQALGWNI--------YASVIALL 182

Qy     170 LIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY----- 224
Db     183 GITMIYTVTGGLAALMYTDTVQTFVILGGACILMGYAFHEVG----GYSGLFDKYLGAAT 238

Qy     225 -----QKPWLGTVDSSEVYSWLDSFLLL---MLGGIPW----------------QAYF 258
Db     239 SLTVSEDPAVGNISSFCYRPRPDSYHLLRHPVTGDLPWPALLLGLTIVSGWYWCSDQVIV 298

Qy     259 QRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKT----TE 314
Db     299 QRCLAGKSLTHIKAGCILCGYLKLTPMFLMVMPGMISRILYPDEVACVVPEVCRRVCGTE 358

Qy     315 E--ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNA 372
Db     359 VGCSNIAYPRLVVKLMPNGLRGLMLAVMLAALMSSLASIFNSSSTLFTMDIY-TRLRPRA 417

Qy     373 SDKEIVWVMRI-TVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQLLCV----LFV 427
Db     418 GDRELLLVGRLWVVFIVVVSVAWLPVVQAAQGGQLFDYIQAVSSYLAPPVSAVFVLALFV 477

Qy     428 KGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMV-- 485
Db     478 PRVNEQGAFWGLIGGLLMGLARLIP----------------------EFSFGSGSCVQP 514

Qy     486 ---TSFLTNICISYLAKYLFE-SGTLPPKLDVFDAVV----------ARHSEENMDKTI 530
Db     515 SACPAFLCGVHYLYFAIVLFFCSGLLTLTVSLCTAPIPRKHLHRLVFSLRHSKE------ 568

Qy     531 LVKNENIKLDE 541
Db     569 --EREDLDADE 577


RESULT 7
US-10-162-012-30
; Sequence 30, Application US/10162012
; Publication No. US20030051660A1
; GENERAL INFORMATION:
;  APPLICANT: Curtis, Rory A.J.
;  APPLICANT:  Silos-Santiago, Inmaculada
;  APPLICANT:  Gu, Wei
;  TITLE OF INVENTION: NOVEL HUMAN ION CHANNEL AND TRANSPORTER FAMILY MEMBERS
;  FILE REFERENCE: 10448-190001
;  CURRENT APPLICATION NUMBER: US/10/162,012
;  CURRENT FILING DATE:  2002-06-04
;  PRIOR APPLICATION NUMBER: US 60/209,845
```

; PRIOR FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: US 09/875,321
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: PCT/US01/18340
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/209,257
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 09/875,423
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: PCT/US01/18398
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/209,238
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 09/875,363
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: PCT/US01/18247
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/227,068
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: US 09/928,530
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/25475
; PRIOR FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: US 60/226,770
; PRIOR FILING DATE: 2000-08-21
; PRIOR APPLICATION NUMBER: US 09/934,421
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: PCT/US01/26096
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/279,281
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 10/109,029
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: PCT/US02/09728
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 60/290,288
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US (not assigned)
; PRIOR FILING DATE: 2002-05-13
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 672
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-162-012-30

Query Match 9.8%; Score 292; DB 9; Length 672;
Best Local Similarity 24.1%; Pred. No. 6.3e-17;
Matches 147; Conservative 91; Mismatches 237; Indels 136; Gaps 22;

```
Qy       8 LIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY 67
           :: |  ::||:: ||:|:  || | |:         : || :     | :: |: :| |:
Db      26 ILVIAAYFLLVIGVGLWSMCRT-NRGTV----GGYFLAGRSMVWWPVGASLFASNIGSGH 80

Qy      68 INGTAEAVYVPGYGLAWAQAPIGYSLS----LILGGLFFAKPMRSKGYVTMLDPFQQIYG 123
             | |      | | |   |  |:  :    ::| |  ||    : | :||     |
Db      81 FVGLA------GTGAASGLAVAGFEWNALFVVLLLGWLFAPVYLTAGVITM----PQYLR 130

Qy     124 KRMGG-----------LLFIPALMGEMFWAAAIF--SALGATISVIIDVDMHISVIISA 169
           || ||              |:|   :    :: |:|   |||  |         : |||
Db     131 KRFGGRRIRLYLSVLSLFLYIFTKISVDMFSGAVFIQQALGWNI--------YASVIALL 182

Qy     170 LIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY----- 224
             |  :||: ||| :: ||| || | |   |  | : :|         |:: :  ||
Db     183 GITMIYTVTGGLAALMYTDTVQTFVILGGACILMGYAFHEVG----GYSGLFDKYLGAAT 238

Qy     225 -----QKPWLGTVDSSEVYSWLDSFLLL---MLGGIPW-----------------QAYF 258
                 : | :| : |       ||: ||   : | :||                   |
Db     239 SLTVSEDPAVGNISSFCYRPRPDSYHLLRHPVTGDLPWPALLLGLTIVSGWYWCSDQVIV 298

Qy     259 QRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKT----TE 314
           || |:  | |: :    |  :  |      :: | |      :: |  :|:       ||
```

```
Db     299 QRCLAGKSLTHIKAGCILCGYLKLTPMFLMVMPGMISRILYPDEVACVVPEVCRRVCGTE 358

Qy     315 E--ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNA 372
             :::  | ::  | |  :    |  : ||:|||  |   |:|::|  :||     |  |
Db     359 VGCSNIAYPRLVVKLMPNGLRGLMLAVMLAALMSSLASIFNSSSTLFTMDIY-TRLRPRA 417

Qy     373 SDKEIVWVMRI-TVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQLLCV----LFV 427
           |:|:: | |:  ||:   |   : ::     |  :     |   : | :  |    |||
Db     418 GDRELLLVGRLWVVFIVVVSVAWLPVVQAAQGGQLFDYIQAVSSYLAPPVSAVFVLALFV 477

Qy     428 KGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMV-- 485
             |  ||  | : || : :    |                          :| | : : |
Db     478 PRVNEQGAFWGLIGGLLMGLARLIP----------------------EFSFGSGSCVQP 514

Qy     486 ---TSFLTNICISYLAKYLFE-SGTLPPKLDVFDAVV----------ARHSEENMDKTI 530
               :||  :   | |  ||  || |  : :  | :           |||:|
Db     515 SACPAFLCGVHYLYFAIVLFFCSGLLTLTVSLCTAPIPRKHLHRLVFSLRHSKE------ 568

Qy     531 LVKNENIKLDE 541
             : |::  ||
Db     569 --EREDLDADE 577
```

RESULT 8
US-10-237-859-2
; Sequence 2, Application US/10237859
; Publication No. US20030022309A1
; GENERAL INFORMATION:
; APPLICANT: MERKULOV, Gennady V. et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001012CON
; CURRENT APPLICATION NUMBER: US/10/237,859
; CURRENT FILING DATE: 2002-09-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Human
US-10-237-859-2

Query Match 9.3%; Score 277.5; DB 9; Length 596;
Best Local Similarity 22.7%; Pred. No. 9.4e-16;
Matches 128; Conservative 91; Mismatches 179; Indels 165; Gaps 22;

```
Qy       5 VEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVG 64
           |  :| | |:: | : |||:: |     ::         : |||:     | :: |:  |
Db      18 VADIIVITVYFALNVAVGIWSSCR-----ASRNTVNGYFLAGRDMTWWPIGASLFASSEG 72

Qy      65 GGYI-----NGTAEAVYVPGY---------GLAWAQAPIGYSLSLILGGLFFAKPMRSKG 110
           |       :| |  : | |:          |||   ||  |  :
Db      73 SGLFIGLAGSGAAGGLAVAGFEWNATYVLLALAWVFVPIYISSEI---------------- 117

Qy     111 YVTMLDPFQQIY-GKRMGGLLFIPALMGEMF-------WAAAIFSALGATISVIIDVDMH 162
             ||: :  |: | |:|:   | : :|:  :|       :| |:|      : : :  : :
Db     118 -VTLPEYIQKRYGGQRIRMYLSVLSLLLSVFTKISLDLYAGALF------VHICLGWNFY 170

Qy     163 ISVIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADI-GFTAVH 221
           :| |::  |  |||: ||| :| ||| :|   : ||  |     |:  |   | |:  :
Db     171 LSTILTLGITALYTIAGGLAAVIYTDALQTLIMVVGAVI-----LTIKAFDQIGGYGQLE 225

Qy     222 AKYQK-------------------------------PWLGTVDSSEVYSWLDSFLLLMLG 250
           | | :                             || [           :| | ::
Db     226 AAYAQAIPSRTIANTTCHLPRTDAMHMFRDPHTGDLPWTGM----------TFGLTIMA 274

Qy     251 GIPW---QAYFQRVLSSSSATYAQVLSFLAAF------GCLVM--------------AI 286
             |   |  || ||:    :|:  | ||::      | ::|              :
Db     275 TWYWCTDQVIVQRSLSARDLNHAKAGSILASYLKMLPMGLIIMPGMISRALFPDDVGCVV 334

Qy     287 PAILIGAIGASTDWNQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMS 346
           |:  : | ||    :  ||   ||   |   ::|| |: |          :  : ||:||
```

```
Db    335 PSECLRACGAEVGCSNIAY----PKLVME---LMPIGLRGLM--------IAVMLAALMS 379

Qy    347 SADSSILSASSMFARNIYQLSFRQNASDKEIVWVMR-ITVFVFGASATAMALLTKTVYGL 405
          |  |   |:|:::|  :|::   |  : ::|:: | | : | : | |   : :|  :  |
Db    380 SLTSIFNSSSTLFTMDIWR-RLRPRSGERELLLVGRLVIVALIGVSVAWIPVLQDSNSGQ 438

Qy    406 WYLSSDLVYIVIFPQLLCV----LFVKGTNTYGAVAGYVSGLFLRITG------------ 449
           ::    |   : | :  |    :| :  |  ||  | ::|| :  |
Db    439 LFIYMQSVTSSLAPPVTAVFVLGVFWRRANEQGAFWGLIAGLVVGATRLVLEFLNPAPPC 498

Qy    450 GEP-----------YLYLQPLIF 461
          |||           ||:    :|
Db    499 GEPDTRPAVLGSIHYLHFAVALF 521


RESULT 9
US-09-740-026A-2
; Sequence 2, Application US/09740026A
; Patent No. US20020081678A1
; GENERAL INFORMATION:
;  APPLICANT: MERKULOV, Gennady V. et al.
;  TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL001012
;  CURRENT APPLICATION NUMBER: US/09/740,026A
;  CURRENT FILING DATE:  2000-12-20
;  NUMBER OF SEQ ID NOS: 101
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-09-740-026A-2

  Query Match             9.3%;  Score 277.5;  DB 10;  Length 596;
  Best Local Similarity   22.7%;  Pred. No. 9.4e-16;
  Matches  128;  Conservative   91;  Mismatches  179;  Indels  165;  Gaps   22;

Qy      5 VEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVG 64
          |  :| | |:: | : |||| :: |     ::       : |||:     | :: |:  |
Db     18 VADIIVITVYFALNVAVGIWSSCR-----ASRNTVNGYFLAGRDMTWWPIGASLFASSEG 72

Qy     65 GGYI-----NGTAEAVYVPGY---------GLAWAQAPIGYSLSLILGGLFFAKPMRSKG 110
          |        :| | :  : | |:          |||   ||  |  :
Db     73 SGLFIGLAGSGAAGGLAVAGFEWNATYVLLALAWVFVPIYISSEI-------------- 117

Qy    111 YVTMLDPFQQIY-GKRMGGLLFIPALMGEMF-------WAAAIFSALGATISVIIDVDMH 162
          ||: :  |: | |:|:   | : :|:  :|       :| |:|      : : :  : :
Db    118 -VTLPEYIQKRYGGQRIRMYLSVLSLLLSVFTKISLDLYAGALF------VHICLGWNFY 170

Qy    163 ISVIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADI-GFTAVH 221
          :| |::  |  |||: ||| :| ||| :|   : ||  |     |:  |   | |:  :
Db    171 LSTILTLGITALYTIAGGLAAVIYTDALQTLIMVVGAVI-----LTIKAFDQIGGYGQLE 225

Qy    222 AKYQK-----------------------------PWLGTVDSSEVYSWLDSFLLLMLG 250
          | | :                             || |            :| | ::
Db    226 AAYAQAIPSRTIANTTCHLPRTDAMHMFRDPHTGDLPWTGM----------TFGLTIMA 274

Qy    251 GIPW---QAYFQRVLSSSSATYAQVLSFLAAF------GCLVM--------------AI 286
             |   |  || ||:    :|:  | ||::      | ::|               :
Db    275 TWYWCTDQVIVQRSLSARDLNHAKAGSILASYLKMLPMGLIIMPGMISRALFPDDVGCVV 334

Qy    287 PAILIGAIGASTDWNQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMS 346
          |:  : | ||    :  ||    ||   |   ::|| |: |          :  : ||:||
Db    335 PSECLRACGAEVGCSNIAY----PKLVME---LMPIGLRGLM--------IAVMLAALMS 379

Qy    347 SADSSILSASSMFARNIYQLSFRQNASDKEIVWVMR-ITVFVFGASATAMALLTKTVYGL 405
          |  |   |:|::|  :|::   |  : ::|:: | | : | : | |   : :|  :  |
Db    380 SLTSIFNSSSTLFTMDIWR-RLRPRSGERELLLVGRLVIVALIGVSVAWIPVLQDSNSGQ 438

Qy    406 WYLSSDLVYIVIFPQLLCV----LFVKGTNTYGAVAGYVSGLFLRITG------------ 449
           ::    |   : | :  |    :| :  |  ||  | ::|| :  |
Db    439 LFIYMQSVTSSLAPPVTAVFVLGVFWRRANEQGAFWGLIAGLVVGATRLVLEFLNPAPPC 498

Qy    450 GEP-----------YLYLQPLIF 461
          |||           ||:    :|
Db    499 GEPDTRPAVLGSIHYLHFAVALF 521


RESULT 10
US-10-119-988-5
; Sequence 5, Application US/10119988
; Publication No. US20030054453A1
; GENERAL INFORMATION:
;  APPLICANT: Curtis, Rory A.J.
;  APPLICANT:  Chen, Hong
;  APPLICANT:  Millennium Pharmaceuticals Inc.
;  TITLE OF INVENTION: 68723, Sodium/Glucose Cotransporter
;  TITLE OF INVENTION:  Family Members and Uses Therefor
;  FILE REFERENCE: MPI01-103P1RNM
;  CURRENT APPLICATION NUMBER: US/10/119,988
;  CURRENT FILING DATE:  2002-04-10
;  PRIOR APPLICATION NUMBER: 60/282,764
;  PRIOR FILING DATE: 2001-04-10
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 643
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-119-988-5

  Query Match             9.3%;  Score 277.5;  DB 9;  Length 643;
  Best Local Similarity   22.7%;  Pred. No. 1e-15;
  Matches  128;  Conservative   91;  Mismatches  179;  Indels  165;  Gaps   22;

Qy      5 VEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVG 64
          |  :| | |:: | : |||| :: |     ::       : |||:     | :: |:  |
Db     65 VADIIVITVYFALNVAVGIWSSCR-----ASRNTVNGYFLAGRDMTWWPIGASLFASSEG 119

Qy     65 GGYI-----NGTAEAVYVPGY---------GLAWAQAPIGYSLSLILGGLFFAKPMRSKG 110
          |        :| | :  : | |:          |||   ||  |  :
Db    120 SGLFIGLAGSGAAGGLAVAGFEWNATYVLLALAWVFVPIYISSEI-------------- 164

Qy    111 YVTMLDPFQQIY-GKRMGGLLFIPALMGEMF-------WAAAIFSALGATISVIIDVDMH 162
          ||: :  |: | |:|:   | : :|:  :|       :| |:|      : : :  : :
Db    165 -VTLPEYIQKRYGGQRIRMYLSVLSLLLSVFTKISLDLYAGALF------VHICLGWNFY 217

Qy    163 ISVIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADI-GFTAVH 221
          :| |::  |  |||: ||| :| ||| :|   : ||  |     |:  |   | |:  :
Db    218 LSTILTLGITALYTIAGGLAAVIYTDALQTLIMVVGAVI-----LTIKAFDQIGGYGQLE 272

Qy    222 AKYQK-----------------------------PWLGTVDSSEVYSWLDSFLLLMLG 250
          | | :                             || |            :| | ::
Db    273 AAYAQAIPSRTIANTTCHLPRTDAMHMFRDPHTGDLPWTGM----------TFGLTIMA 321

Qy    251 GIPW---QAYFQRVLSSSSATYAQVLSFLAAF------GCLVM--------------AI 286
             |   |  || ||:    :|:  | ||::      | ::|               :
Db    322 TWYWCTDQVIVQRSLSARDLNHAKAGSILASYLKMLPMGLIIMPGMISRALFPDDVGCVV 381

Qy    287 PAILIGAIGASTDWNQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAAVMS 346
          |:  : | ||    :  ||    ||   |   ::|| |: |          :  : ||:||
Db    382 PSECLRACGAEVGCSNIAY----PKLVME---LMPIGLRGLM--------IAVMLAALMS 426

Qy    347 SADSSILSASSMFARNIYQLSFRQNASDKEIVWVMR-ITVFVFGASATAMALLTKTVYGL 405
          |  |   |:|::|  :|::   |  : ::|:: | | : | : | |   : :|  :  |
Db    427 SLTSIFNSSSTLFTMDIWR-RLRPRSGERELLLVGRLVIVALIGVSVAWIPVLQDSNSGQ 485

Qy    406 WYLSSDLVYIVIFPQLLCV----LFVKGTNTYGAVAGYVSGLFLRITG------------ 449
           ::    |   : | :  |    :| :  |  ||  | ::|| :  |
Db    486 LFIYMQSVTSSLAPPVTAVFVLGVFWRRANEQGAFWGLIAGLVVGATRLVLEFLNPAPPC 545

Qy    450 GEP-----------YLYLQPLIF 461
          |||           ||:    :|
Db    546 GEPDTRPAVLGSIHYLHFAVALF 568
```

```
RESULT 11
US-10-237-859-4
; Sequence 4, Application US/10237859
; Publication No. US20030022309A1
; GENERAL INFORMATION:
;  APPLICANT: MERKULOV, Gennady V. et al
;  TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL001012CON
;  CURRENT APPLICATION NUMBER: US/10/237,859
;  CURRENT FILING DATE:  2002-09-10
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Oryctolagus cuniculus
US-10-237-859-4

  Query Match             9.3%;  Score 277;  DB 9;  Length 596;
  Best Local Similarity   22.5%;  Pred. No. 1e-15;
  Matches  119;  Conservative   92;  Mismatches  171;  Indels  148;  Gaps   20;

Qy       5 VEGLIAIIVFYLLILLVGIWAAWR-TKNSGSAEERSEAIIVGGRDI---------- 49
Db      18 VTDIVVITVYFALNVAVGIWSSCRASRNTVS------GYFLAGRDMTWWPIGASLFGSSE 71

Qy      50 --GLLVGGFTMTATWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMR 107
Db      72 GSGLFIG-------LAGSGAAGGLAVA------GFDWNATYVLLALAWVFGAIYI----- 113

Qy     108 SKGYVTMLDPFQQIY-GKRMGGLLFIPALMGEMF-------WAAAIFSALGATISVIIDV 159
Db     114 SSEIVTLAEYIQKRFGGQRIRMYLSVLSLLLSVFTKISLDLYAGALF------VHICLGW 167

Qy     160 DMHISVIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADI-GFT 218
Db     168 NFYLSTILTLTITALYTITGGLVAVIYTDALQTLIMVVGAVI-----LAIKAFHQIDGYG 222

Qy     219 AVHAKYQK-------------------------------PWLGTVDSSEVYSWLDSFLLL 247
Db     223 QMEAAYARAIPSRTVANTTCHLPRADAMHMFRDPYTGDLPWTGM----------TFGLT 271

Qy     248 MLGGIPW---QAYFQRVLSSSSATYAQVLSFLAAF------GCLVM---------- 284
Db     272 IMATWYWCTDQVIVQRSLSARNLNHAKAGSILASYLKMLPMGLMIMPGMISRALFPDEVG 331

Qy     285 -AIPAILIGAIGASTDWNQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAA 343
Db     332 CVVPSECLRACGAEIGCSNIAY--------------PKLVMELMPVGLRGLMIAVMMPA 376

Qy     344 VMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITVFV-FGASATAMALLTKTV 402
Db     377 LMSSLSSIFNSSSTLFTMDIWR-RLRPCASERELLLVGRLVIVVLIGVSVAWIPVLQGSN 435

Qy     403 YGLWYLSSDLVYIVIFPQLLCV----LFVKGTNTYGAVAGYVSGLFLRIT 448
Db     436 GGQLFIYMQSVTSSLAPPVTAVFTLGIFWQRANEQGAFWGLLAGLAVGAT 485


RESULT 12
US-09-740-026A-4
; Sequence 4, Application US/09740026A
; Patent No. US20020081678A1
; GENERAL INFORMATION:
;  APPLICANT: MERKULOV, Gennady V. et al.
;  TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL001012
;  CURRENT APPLICATION NUMBER: US/09/740,026A
;  CURRENT FILING DATE:  2000-12-20
;  NUMBER OF SEQ ID NOS: 101
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Oryctolagus cuniculus
US-09-740-026A-4

  Query Match             9.3%;  Score 277;  DB 10;  Length 596;
  Best Local Similarity   22.5%;  Pred. No. 1e-15;
  Matches  119;  Conservative   92;  Mismatches  171;  Indels  148;  Gaps   20;

Qy       5 VEGLIAIIVFYLLILLVGIWAAWR-TKNSGSAEERSEAIIVGGRDI---------- 49
Db      18 VTDIVVITVYFALNVAVGIWSSCRASRNTVS------GYFLAGRDMTWWPIGASLFGSSE 71

Qy      50 --GLLVGGFTMTATWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMR 107
Db      72 GSGLFIG-------LAGSGAAGGLAVA------GFDWNATYVLLALAWVFGAIYI----- 113

Qy     108 SKGYVTMLDPFQQIY-GKRMGGLLFIPALMGEMF-------WAAAIFSALGATISVIIDV 159
Db     114 SSEIVTLAEYIQKRFGGQRIRMYLSVLSLLLSVFTKISLDLYAGALF------VHICLGW 167

Qy     160 DMHISVIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADI-GFT 218
Db     168 NFYLSTILTLTITALYTITGGLVAVIYTDALQTLIMVVGAVI-----LAIKAFHQIDGYG 222

Qy     219 AVHAKYQK-------------------------------PWLGTVDSSEVYSWLDSFLLL 247
Db     223 QMEAAYARAIPSRTVANTTCHLPRADAMHMFRDPYTGDLPWTGM----------TFGLT 271

Qy     248 MLGGIPW---QAYFQRVLSSSSATYAQVLSFLAAF------GCLVM---------- 284
Db     272 IMATWYWCTDQVIVQRSLSARNLNHAKAGSILASYLKMLPMGLMIMPGMISRALFPDEVG 331

Qy     285 -AIPAILIGAIGASTDWNQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAVSAA 343
Db     332 CVVPSECLRACGAEIGCSNIAY--------------PKLVMELMPVGLRGLMIAVMMPA 376

Qy     344 VMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITVFV-FGASATAMALLTKTV 402
Db     377 LMSSLSSIFNSSSTLFTMDIWR-RLRPCASERELLLVGRLVIVVLIGVSVAWIPVLQGSN 435

Qy     403 YGLWYLSSDLVYIVIFPQLLCV----LFVKGTNTYGAVAGYVSGLFLRIT 448
Db     436 GGQLFIYMQSVTSSLAPPVTAVFTLGIFWQRANEQGAFWGLLAGLAVGAT 485


RESULT 13
US-09-738-626-6949
; Sequence 6949, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
```

; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6949
; LENGTH: 524
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6949

Query Match 9.2%; Score 272.5; DB 9; Length 524;
Best Local Similarity 21.7%; Pred. No. 2.1e-15;
Matches 123; Conservative 104; Mismatches 230; Indels 111; Gaps 21;

Qy 8 LIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY 67
Db 8 IIAIVIYMLVMVLIGYWSYRKT-------EKYDDYMLAGRGLNPFVAAMSAGASDMSGWL 60

Qy 68 INGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMR-----SKGYVTMLDPFQQIY 122
Db 61 LMGLPGALFVTGMSELW--IAVGLTIGAWANWMWVAPRLRSYSEISANSITLPSFFENRL 118

Qy 123 GKRMGGLLFIPALMGEMFWAAAIFSAL---GATISVIIDVDMHISVIISALIATLYTLVG 179
Db 119 RDKSRALRIIAALIIIVFFTFYISSGMVAGGVYWESTFGGDYLLGMAIVAGVTVLYTFIG 178

Qy 180 GLYSVAYTDVVQLFCIFVGLWISVP----FALSHPAVADIGFTAVHAKYQKPWLGTVDSS 235
Db 179 GFLAVSYTDAVQGTIMFFSL-IIVPVMAYFALANP--MDI-WSFANSNDYGPHTDGIGNP 234

Qy 236 EVYSWLDSF-LLLMLGGIPW----------QAYFQRVLSSSSATYAQVLSFLAAFGCLVM 284
Db 235 TYFSMISGISAAAIIGNLGWGLGYFGQPHIVVRFMALRTPAEAKQGRRIGISWMIICLIG 294

Qy 285 AIPAILIGAIGASTDWNQTAYGLPDPKTTEE---ADMILPIVLQYLCPVYISFFGLGAVS 341
Db 295 ATFTAIISTVFFAQN--------PDANITDTRAYESIFLDLARMLFHPL-IAGLILTAVL 345

Qy 342 AAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKT 401
Db 346 AAIMSTMSSQLLVTASSLIEDLLKVVKKDSLSERTLIMLSRATVIILAIIAAAMAINP-- 403

Qy 402 VYGLWYLSSDLVYIVIF----------PQLLCVLFVKGTNTYGAVAGYVSGLFLRITGGE 451
Db 404 -------SDSILGLVGFAWAGFGSAFGPIILASLYWKRLNAAGAISGMITGAIVSIAWG- 455

Qy 452 PYLYLQPLI-----FYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGT 506
Db 456 ----MSPLSDTLYEIIPG-----------FALATIVMVV-------VSLLTK------- 485

Qy 507 LPPKLDVF-------DAVVARHSEENMD 527
Db 486 -EPSEEILNEFETAKDLAAAVESNEDVD 512

RESULT 14
US-10-119-988-2
; Sequence 2, Application US/10119988
; Publication No. US20030054453A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A.J.
; APPLICANT: Chen, Hong
; APPLICANT: Millennium Pharmaceuticals Inc.
; TITLE OF INVENTION: 68723, Sodium/Glucose Cotransporter
; TITLE OF INVENTION: Family Members and Uses Therefor
; FILE REFERENCE: MPI01-103P1RNM
; CURRENT APPLICATION NUMBER: US/10/119,988
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 60/282,764
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 2
; LENGTH: 664
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-119-988-2

Query Match 9.1%; Score 269.5; DB 9; Length 664;
Best Local Similarity 22.1%; Pred. No. 5.3e-15;
Matches 128; Conservative 91; Mismatches 179; Indels 181; Gaps 22;

Qy 5 VEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVG 64
Db 65 VADIIVITVYFALNVAVGIWSSCR-----ASRNTVNGYFLAGRDMTWWPIGASLFASSEG 119

Qy 65 GGYI-----NGTAEAVYVPGY---------GLAWAQAPIGYSLSLILGGLFFAKPMRSKG 110
Db 120 SGLFIGLAGSGAAGGLAVAGFEWNATYVLLALAWVFVPIYISSEI--------------- 164

Qy 111 YVTMLDPFQQIY-GKRMGGLLFIPALMGEMF-------WAAAIFSALGATISVIIDVDMH 162
Db 165 -VTLPEYIQKRYGGQRIRMYLSVLSLLLSVFTKISLDLYAGALF------VHICLGWNFY 217

Qy 163 ISVIISALIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADI-GFTAVH 221
Db 218 LSTILTLGITALYTIAGGLAAVIYTDALQTLIMVVGAVI-----LTIKAFDQIGGYGQLE 272

Qy 222 AKYQK-------------------------------PWLGTVDSSEVYSWLDSFLLLMLG 250
Db 273 AAYAQAIPSRTIANTTCHLPRTDAMHMFRDPHTGDLPWTGM----------TFGLTIMA 321

Qy 251 GIPW---QAYFQRVLSSSSATYAQVLSFLAAF------GCLVM---------------- 284
Db 322 TWYWCTDQVIVQRSLSARDLNHAKAGSILASYLKMLPMGLIIMPGMISRALFPGAHVYEE 381

Qy 285 --------------AIPAILIGAIGASTDWNQTAYGLPDPKTTEEADMILPIVLQYLCPV 330
Db 382 RHQVSVSRTDDVGCVVPSECLRACGAEVGCSNIAY----PKLVME---LMPIGLRGLM-- 432

Qy 331 YISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMR-ITVFVFG 389
Db 433 ------IAVMLAALMSSLTSIFNSSSTLFTMDIWR-RLRPRSGERELLLVGRLVIVALIG 485

Qy 390 ASATAMALLTKTVYGLWYLSSDLVYIVIFPQLLCV----LFVKGTNTYGAVAGYVSGLFL 445
Db 486 VSVAWIPVLQDSNSGQLFIYMQSVTSSLAPPVTAVFVLGVFWRRANEQGAFWGLIAGLVV 545

Qy 446 RITG------------GEP-----------YLYLQPLIF 461
Db 546 GATRLVLEFLNPAPPCGEPDTRPAVLGSIHYLHFAVALF 584

RESULT 15
US-10-173-123-9
; Sequence 9, Application US/10173123
; Publication No. US20030027301A1
; GENERAL INFORMATION:
; APPLICANT: Hu, Yi
; APPLICANT: Nepomnichy, Boris
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Mathur, Brian
; APPLICANT: Friddle, Carl Johan
; TITLE OF INVENTION: No. US20030027301A1el Human Transporter Proteins and
; TITLE OF INVENTION: Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0358-USA
; CURRENT APPLICATION NUMBER: US/10/173,123
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: US 60/298,241
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 674
; TYPE: PRT

; ORGANISM: homo sapiens
US-10-173-123-9

Query Match 8.7%; Score 260; DB 9; Length 674;
Best Local Similarity 21.2%; Pred. No. 3.6e-14;
Matches 113; Conservative 87; Mismatches 177; Indels 156; Gaps 19;

```
Qy       1 MAFHVEGLIAIIVFYLLILLVGIWAA-------------------WRTKNSGSAEERSE 40
Db      24 VGLHAYDISVVVIYFVFVIAVGIWSSIRASRGTIGGYFLAGRSMSWWPIGASLMSSNVGS 83

Qy      41 AIIVG----GRDIGLLVGGFTMTATWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLI 96
Db      84 GLFIGLAGTGAAGGLAVGGFEWNATWL-----------------------------LL 112

Qy      97 LGGLFFAKPMRSKGYVTMLDPFQQIYGKRMGG-----------LLFIPALMGEMFWAAA 144
Db     113 ALGWVFVPVYIAAGVVTM----PQYLKKRFGGQRIQVYMSVLSLILYIFTKISTDIFSGA 168

Qy     145 IFSALGATISVIIDVDMHISVIISALIATLYTLVGGLYSVAYTDVVQ--------LFCIF 196
Db     169 LF------IQMALGWNLYLSTGILLVVTAVYTIAGGLMAVIYTDALQTVIMVGGALVLMF 222

Qy     197 VGL----W-------------ISVPFALSH-----------PAVADIGFTAVHAKYQK 226
Db     223 LGFQDVGWYPGLEQRYRQAIPNVTVPNTTCHLPRPDAFHMLRDPVSGDI---------- 271

Qy     227 PWLGTVDSSEVYSWLDSFLLLMLGGIPW---QAYFQRVLSSSSATYAQVLSFLAAFGCLV 283
Db     272 PWPGLI----------FGLTVLATWCWCTDQVIVQRSLSAKSLSHAKGGSVLGGYLKIL 320

Qy     284 MAIPAILIGAIGASTDWNQTAYGLPDPKTTEE--------ADMILPIVLQYLCPVYISFF 335
Db     321 PMFFIVMPGMISRALFPDEV--GCVDPDVCQRICGARVGCSNIAYPKLVMALMPVGLRGL 378

Qy     336 GLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRI-TVFVFGASATA 394
Db     379 MIAVIMAALMSSLTSIFNSSSTLFTIDVWQ-RFRRKSTEQELMVVGRVFVVFLVVISILW 437

Qy     395 MALLTKTVYG--LWYLSSDLVYIV--IFPQLLCVLFVKGTNTYGAVAGYVSGL 443
Db     438 IPIIQSSNSGQLFDYIQAVTSYLAPPITALFLLAIFCKRVTEPGAFWGLVFGL 490
```

Search completed: March 31, 2003, 18:48:39
Job time : 595 secs

OM protein - protein search, using sw model

Run on: March 31, 2003, 18:34:36 ; Search time 39 Seconds
(without alignments)
437.572 Million cell updates/sec

Title: US-10-069-541-6
Perfect score: 2972
Sequence: 1 MAFHVEGLIAIIVFYLLILL.....EAFLDVDSSPEGSGTEDNLQ 580

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	308.5	10.4	662	1	US-07-841-651-4	Sequence 4, Appli
2	298	10.0	672	1	US-07-841-651-2	Sequence 2, Appli
3	298	10.0	672	1	US-07-841-651-3	Sequence 3, Appli
4	262.5	8.8	518	4	US-09-134-001C-4744	Sequence 4744, Ap
5	218	7.3	618	4	US-08-595-553A-2	Sequence 2, Appli
6	206.5	6.9	530	4	US-09-134-001C-4510	Sequence 4510, Ap
7	200	6.7	635	2	US-09-014-969-11	Sequence 11, Appl
8	138	4.6	482	4	US-09-134-001C-4309	Sequence 4309, Ap
9	123	4.1	496	4	US-09-134-001C-3001	Sequence 3001, Ap
10	119	4.0	499	4	US-09-134-001C-5370	Sequence 5370, Ap
11	116	3.9	832	2	US-08-677-734A-12	Sequence 12, Appl
12	116	3.9	832	4	US-09-097-053-12	Sequence 12, Appl
13	113.5	3.8	503	4	US-09-068-195-24	Sequence 24, Appl
14	110	3.7	599	1	US-08-295-814E-11	Sequence 11, Appl
15	110	3.7	599	1	US-08-240-783B-4	Sequence 4, Appli
16	110	3.7	599	3	US-09-084-813-4	Sequence 4, Appli
17	110	3.7	599	4	US-09-343-361-11	Sequence 11, Appl
18	110	3.7	599	5	PCT-US92-09662-4	Sequence 4, Appli
19	109.5	3.7	405	4	US-09-134-001C-4999	Sequence 4999, Ap
20	109	3.7	367	4	US-09-134-001C-5557	Sequence 5557, Ap
21	108	3.6	423	4	US-09-134-001C-3599	Sequence 3599, Ap
22	108	3.6	599	1	US-07-879-617A-10	Sequence 10, Appl
23	108	3.6	599	1	US-08-753-985-10	Sequence 10, Appl
24	106	3.6	831	2	US-08-677-734A-11	Sequence 11, Appl
25	106	3.6	831	4	US-09-097-053-11	Sequence 11, Appl
26	105.5	3.5	599	1	US-08-301-722A-5	Sequence 5, Appli
27	104.5	3.5	416	4	US-09-333-208-2	Sequence 2, Appli
28	104.5	3.5	416	4	US-09-333-254-2	Sequence 2, Appli
29	104.5	3.5	416	4	US-09-183-270-2	Sequence 2, Appli
30	104.5	3.5	834	2	US-08-677-734A-9	Sequence 9, Appli
31	104.5	3.5	834	2	US-08-677-734A-10	Sequence 10, Appl
32	104.5	3.5	834	4	US-09-097-053-9	Sequence 9, Appli
33	104.5	3.5	834	4	US-09-097-053-10	Sequence 10, Appl
34	102.5	3.4	371	2	US-08-928-692-20	Sequence 20, Appl
35	102.5	3.4	371	4	US-09-339-972-20	Sequence 20, Appl
36	101.5	3.4	408	4	US-09-134-001C-4130	Sequence 4130, Ap
37	101.5	3.4	1584	4	US-09-251-645-6	Sequence 6, Appli
38	101	3.4	478	4	US-09-134-001C-4637	Sequence 4637, Ap
39	101	3.4	1296	4	US-09-462-136-9	Sequence 9, Appli
40	100	3.4	413	4	US-09-134-001C-3702	Sequence 3702, Ap
41	100	3.4	490	4	US-09-134-001C-5116	Sequence 5116, Ap
42	99.5	3.3	437	4	US-09-134-001C-4808	Sequence 4808, Ap
43	99.5	3.3	473	4	US-09-637-118B-2	Sequence 2, Appli
44	99.5	3.3	522	4	US-09-142-732-2	Sequence 2, Appli
45	99.5	3.3	635	1	US-07-879-617A-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1
US-07-841-651-4
; Sequence 4, Application US/07841651
; Patent No. 5410031
; GENERAL INFORMATION:
; APPLICANT: Pajor, Ana M
; APPLICANT: Wright, Ernest M
; TITLE OF INVENTION: Cloning and Functional Expression of a
; TITLE OF INVENTION: Mammalian Na+/Nucleoside Cotransporter: A Member of the
; TITLE OF INVENTION: SGLT Family
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheldon & Mak
; STREET: 225 South Lake Avenue, Ninth Floor
; CITY: Pasadena
; STATE: California
; COUNTRY: USA
; ZIP: 91101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/841,651
; FILING DATE: 19920224
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandel, SaraLynn
; REGISTRATION NUMBER: 31,853
; REFERENCE/DOCKET NUMBER: 8772
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (818) 796-4000
; TELEFAX: (818) 795-6321
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 662 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Oryctolagus cuniculus
US-07-841-651-4

Query Match 10.4%; Score 308.5; DB 1; Length 662;
Best Local Similarity 23.4%; Pred. No. 1.9e-21;
Matches 154; Conservative 110; Mismatches 238; Indels 155; Gaps 26;

Qy 11 IIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYING 70

```
Db      32 IVIYFLVVMAVGLWAMFST-NRGTV----GGFFLAGRSMVWWPIGASLFASNIGSGHFVG 86

Qy      71 TAEAVYVPGYGLAWAQAPIGYS-----LSLILGGLFFAKPMRSKGYVTMLDPFQQIY-GK 124

Db      87 LA------GTGAASGIATGGFEWNALIMVVVLGWVFVPIYIRA-GVVTMPEYLQKRFGGK 139

Qy     125 RMGGLLFIPALMGEMFW--AAAIFSALGAT-ISVIIDVDMHISVIISALIATLYTLVGGL 181

Db     140 RIQIYLSILSLLLYIFTKISADIFS--GAIFIQLTLGLDIYVAIIILLVITGLYTITGGL 197

Qy     182 YSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY----------------Q 225

Db     198 AAVIYTDTLQTAIMMVGSVILTGFAFHEVG----GYEAFTEKYMRAIPSQISYGNTSIPQ 253

Qy     226 KPWLGTVDSSEVYSWLDSFLLLMLGGIPW-----------------QAYFQRVLSSSSA 267

Db     254 KCYTPREDAFHI------FRDAITGDIPWPGLVFGMSILTLWYWCTDQVIVQRCLSAKNL 307

Qy     268 TYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDP------KTTEEADMILP 321

Db     308 SHVKAGCILCGYLKVMPMFLIVMMGMVSRILYTDKVACVVPSECERYCGTRVGCTNIAFP 367

Qy     322 IVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVWVM 381

Db     368 TLVVELMPNGLRGLMLSVMMASLMSSLTSIFNSASTLFTMDIY-TKIRKKASEKELMIAG 426

Qy     382 RI-TVFVFGASATAMALLTKTVYG--LWYLSSDLVYI--VIFPQLLCVLFVKGTNTYGAV 436

Db     427 RLFMLFLIGISIAWVPIVQSAQSGQLFDYIQSITSYLGPPIAAVFLLAIFWKRVNEPGAF 486

Qy     437 AGYVSGLFLRI----------TG------------GEPYLYLQPLIFYPGYYPDDNGIY 473

Db     487 WGLVLGFLIGISRMITEFAYGTGSCMEPSNCPTIICGVHYLYFAIILF------------ 534

Qy     474 NQKFPFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDAVVA-RHSEENMDKTILV 532

Db     535 ----------VISIITVVVVSLFTKPI-------PDVHLYRLCWSLRNSKE-------- 568

Qy     533 KNENIKLD--ELALVKPRQSMTLSSTFTNKEAF----------LDVDSSPEGSGTED 577

Db     569 --ERIDLDAGEEDIQEAPEEATDTEVPKKKKGFFRRAYDLFCGLDQDKGPKMTKEEE 623


RESULT 2
US-07-841-651-2
; Sequence 2, Application US/07841651
; Patent No. 5410031
;  GENERAL INFORMATION:
;    APPLICANT:  Pajor, Ana M
;    APPLICANT:  Wright, Ernest M
;    TITLE OF INVENTION:  Cloning and Functional Expression of a
;    TITLE OF INVENTION:  Mammalian Na+/Nucleoside Cotransporter: A Member of the
;    TITLE OF INVENTION:  SGLT Family
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheldon & Mak
;      STREET:  225 South Lake Avenue, Ninth Floor
;      CITY:  Pasadena
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  91101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/841,651
;      FILING DATE:  19920224
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandel, SaraLynn
;      REGISTRATION NUMBER:  31,853
;      REFERENCE/DOCKET NUMBER:  8772
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (818) 796-4000
;      TELEFAX:  (818) 795-6321
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  672 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-841-651-2

  Query Match             10.0%;  Score 298;  DB 1;  Length 672;
  Best Local Similarity   25.0%;  Pred. No. 2.1e-20;
  Matches  153;  Conservative   89;  Mismatches  232;  Indels  138;  Gaps   25;

Qy       9 IAII-VFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY 67

Db      26 IAVIAAYFLLVIGVGLWSMCRT-NRGTV----GGYFLAGRSMVWWPVGASLFASNIGSGH 80

Qy      68 INGTAEAVYVPGYGLAWAQAPIGYSLS----LILGGLFFAKPMRSKGYVTMLDPFQQIYG 123

Db      81 FVGLA------GTGAANGLAVAGFEWNALFVVLLLGWLFAPVYLTAGVITM----PQYLR 130

Qy     124 KRMGG------------LLFIPALMGEMFWAAAIF--SALGATISVIIDVDMHISVIISA 169

Db     131 KRFGGHRIRLYLSVLSLFLYIFTKISVDMFSGAVFIQQALGWNI--------YASVIALL 182

Qy     170 LIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY----- 224

Db     183 GITMVYTVTGGLAALMYTDTVQTFVIIAGAFILTGYAFHEVG----GYSGLFDKYMGAMT 238

Qy     225 -----QKPWLGTVDSSEVYSWLDSFLLL---MLGGIPW-----------------QAYF 258

Db     239 SLTVSEDPAVGNISSSCYRPRPDSYHLLRDPVTGDLPWPALLLGLTIVSGWYWCSDQVIV 298

Qy     259 QRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKT----TE 314

Db     299 QRCLAGRNLTHIKAGCILCGYLKLTPMFLMVMPGMISRILYPDEVACVAPEVCKRVCGTE 358

Qy     315 E--ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNA 372

Db     359 VGCSNIAYPRLVVKLMPNGLRGLMLAVMLAALMSSLASIFNSSSTLFTMDIYTL--RPRA 416

Qy     373 SDKEIVWVMRITVFVFGASATAMALLTKTVYG---LWYLSSDLVYIV--IFPQLLCVLFV 427

Db     417 GEGELLLVGRLWVVFIVAVSVAWLPVVQAAQGGQLFDYIQSVSSYLAPPVSAVFVVALFV 476

Qy     428 KGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMV-- 485

Db     477 PRVNEKGAFWGLIGGLLMGLARLIP----------------------EFSFGTGSCVRP 513

Qy     486 ---TSFLTNICISYLAKYLFE-SG---------TLP-PKLDVFDAVVA-RHSEENMDKTI 530

Db     514 SACPAFLCRVHYLYFAIVLFFCSGLLIIIVSLCTAPIPRKHLHRLVFSLRHSKE------ 567

Qy     531 LVKNENIKLDEL 542

Db     568 --EREDLDADEL 577


RESULT 3
US-07-841-651-3
; Sequence 3, Application US/07841651
; Patent No. 5410031
;  GENERAL INFORMATION:
;    APPLICANT:  Pajor, Ana M
;    APPLICANT:  Wright, Ernest M
;    TITLE OF INVENTION:  Cloning and Functional Expression of a
;    TITLE OF INVENTION:  Mammalian Na+/Nucleoside Cotransporter: A Member of the
;    TITLE OF INVENTION:  SGLT Family
;    NUMBER OF SEQUENCES:  4
```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheldon & Mak
; STREET: 225 South Lake Avenue, Ninth Floor
; CITY: Pasadena
; STATE: California
; COUNTRY: USA
; ZIP: 91101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/841,651
; FILING DATE: 19920224
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandel, SaraLynn
; REGISTRATION NUMBER: 31,853
; REFERENCE/DOCKET NUMBER: 8772
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (818) 796-4000
; TELEFAX: (818) 795-6321
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 672 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Oryctolagus cuniculus
US-07-841-651-3

Query Match 10.0%; Score 298; DB 1; Length 672;
Best Local Similarity 25.0%; Pred. No. 2.1e-20;
Matches 153; Conservative 89; Mismatches 232; Indels 138; Gaps 25;

```
Qy      9 IAII-VFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGY 67
          ||:|  ::||:: ||:|:  || | |:         : || :     | :: |: :| |:
Db     26 IAVIAAYFLLVIGVGLWSMCRT-NRGTV----GGYFLAGRSMVWWPVGASLFASNIGSGH 80

Qy     68 INGTAEAVYVPGYGLAWAQAPIGYSLS----LILGGLFFAKPMRSKGYVTMLDPFQQIYG 123
            | |      | | |    |  |:  :    ::| |  ||    : | :||     |
Db     81 FVGLA------GTGAANGLAVAGFEWNALFVVLLLGWLFAPVYLTAGVITM----PQYLR 130

Qy    124 KRMGG------------LLFIPALMGEMFWAAAIF--SALGATISVIIDVDMHISVIISA 169
          || ||              |:|   :    :: |:|   |||  |         : |||
Db    131 KRFGGHRIRLYLSVLSLFLYIFTKISVDMFSGAVFIQQALGWNI--------YASVIALL 182

Qy    170 LIATLYTLVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKY----- 224
            |  :||: ||| :: ||| || | |  | :|   :|          |:: :  ||
Db    183 GITMVYTVTGGLAALMYTDTVQTFVIIAGAFILTGYAFHEVG----GYSGLFDKYMGAMT 238

Qy    225 -----QKPWLGTVDSSEVYSWLDSFLLL---MLGGIPW----------------QAYF 258
                : | :| : ||      ||: ||   : | :||                  |
Db    239 SLTVSEDPAVGNISSSCYRPRPDSYHLLRDPVTGDLPWPALLLGLTIVSGWYWCSDQVIV 298

Qy    259 QRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKT----TE 314
          || |:  : |: :    |  :  |      :: | |      :: |   |:       ||
Db    299 QRCLAGRNLTHIKAGCILCGYLKLTPMFLMVMPGMISRILYPDEVACVAPEVCKRVCGTE 358

Qy    315 E--ADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNA 372
          :::  | ::  | |  :    |  : ||:|||  |   |:|::|  :|| |  |  |
Db    359 VGCSNIAYPRLVVKLMPNGLRGLMLAVMLAALMSSLASIFNSSSTLFTMDIYTL--RPRA 416

Qy    373 SDKEIVWVMRITVFVFGASATAMALLTKTVYG---LWYLSSDLVYIV--IFPQLLCVLFV 427
          : |:: | |: |    | : |   : :   |   |: |   |:   :   :  |||
Db    417 GEGELLLVGRLWVVFIVAVSVAWLPVVQAAQGGQLFDYIQSVSSYLAPPVSAVFVVALFV 476

Qy    428 KGTNTYGAVAGYVSGLFLRITGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMV-- 485
            | ||  | : || : :    |                        :| | | : |
Db    477 PRVNEKGAFWGLIGGLLMGLARLIP----------------------EFSFGTGSCVRP 513
```

```
Qy    486 ---TSFLTNICISYLAKYLFE-SG---------TLP-PKLDVFDAVVA-RHSEENMDKTI 530
              :||  :   | |  ||  ||          | | |:  :    | : |||:|
Db    514 SACPAFLCRVHYLYFAIVLFFCSGLLIIIVSLCTAPIPRKHLHRLVFSLRHSKE------ 567

Qy    531 LVKNENIKLDEL 542
            : |::  |||
Db    568 --EREDLDADEL 577
```

RESULT 4
US-09-134-001C-4744
; Sequence 4744, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4744
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4744

Query Match 8.8%; Score 262.5; DB 4; Length 518;
Best Local Similarity 22.2%; Pred. No. 4.2e-17;
Matches 126; Conservative 102; Mismatches 223; Indels 117; Gaps 25;

```
Qy      9 IAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYI 68
          : |||:::::|:::| :  :|      :|      ::||| ||  :   :  |: : |  |
Db     27 VMIIVYFIILLIIGFY-GYRQATGNLSE-----FMLGGRSIGPYITALSAGASDMSGWMI 80

Qy     69 NGTAEAVYVPGYGLAWAQAPIGYSLSLILGGL--FFAKPMRSKGY-------VTMLDPFQ 119
            |   :||  |    |       :: | ||    :|    | : |       :|: | |:
Db     81 MGLPGSVYSTGLSAIW------ITIGLTLGAYINYFVVAPRLRVYTEIAGDAITLPDFFK 134

Qy    120 QIYGKRMGGLLFIPALMGEMFWAAAIFSAL---GATISVIIDVDMHISVIISALIATLYT 176
            :   :  |  |:  :|:      |      |           :: |  ::| |:|   ||
Db    135 NRLDDKKNIIKIISGLIIVVFFTLYTHSGFVSGGKLFESAFGLNYHAGLLIVAIIVIFYT 194

Qy    177 LVGGLYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADI-GFTAVHAKYQ-KPW------ 228
            ||  :|: ||  |   : : : : ||      |:  : |:   |   | ||
Db    195 FFGGYLAVSITDFFQGVIMLIAM-VMVPIV----ALLKLNGWDTFHDIAQMKPTNLDLFR 249

Qy    229 ----LGTVDSSEVYSWLDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVM 284
              || |    ::||          |: :      ::   |    ::|      |   |
Db    250 GTTVLGIV---SLFSW----------GLGYFGQPHIIVRFMSIKSHKLLPKARRLGISWM 296

Qy    285 AIPAILIGAIGASTDW----NQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFGLGAV 340
          |:   |:|||          ::    | ||:|      :  :: | |    :  | | |:
Db    297 AVG--LLGAIGVGLTGISFISERHIKLEDPET------LFIVMSQILFHPLVGGFLLAAI 348

Qy    341 SAAVMSSADSSILSASSMFARNIYQL---SFRQNASDKEIVWVMRITVFVFGASATAMAL 397
            ||:||:  | :|  ||    : |:|   | : ::  || | : |::| :    |  :|
Db    349 LAAIMSTISSQLLVTSSSLTEDFYKLIRGSDKASSHQKEFVLIGRLSVLLVAIVAITIA- 407

Qy    398 LTKTVYGLWYLSSDLVYIV----------IFPQLLCVLFVKGTNTYGAVAGYVSGLFLRI 447
                  |: :  :: :|           | :|  |: |     ||::| |:|  : |
Db    408 --------WHPNDTILNLVGNAWAGFGAAFSPLVLYSLYWKDLTRAGAISGMVAGAVVVI 459

Qy    448 TGGEPYLYLQPLIFYPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFESGTL 507
                :::||     ::    |:|          ::  |: :: |:|:   |    |
Db    460 VW---ISWIKPLATINAFF----GMYE---------IIPGFIVSVLITYIVSKL----TK 499

Qy    508 PPKLDVFDAVVARHSEENMDKTILVKNE 535
```

```
Db    500 KPD----DYVI-----ENLNKVKHVVKE 518

RESULT 5
US-08-595-553A-2
; Sequence 2, Application US/08595553A
; Patent No. 6391579
;  GENERAL INFORMATION:
;    APPLICANT:  NANCY CARRASCO, ET AL.
;    TITLE OF INVENTION:  THYROID SODIUM/IODIDE SYMPORTER AND
;    TITLE OF INVENTION:  NUCLEIC ACID ENCODING SAME
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  AMSTER, ROTHSTEIN & EBENSTEIN
;      STREET:  90 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  U.S.A.
;      ZIP:  10016
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH 1.44 Mb STORAGE
;      MEDIUM TYPE:  DISKETTE
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/595,553A
;      FILING DATE:  FEBRUARY 1, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CRAIG J. ARNOLD
;      REGISTRATION NUMBER:  34,287
;      REFERENCE/DOCKET NUMBER:  96700/393
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 697-5995
;      TELEFAX:  (212) 286-0854 or 286-0082
;      TELEX:  TWX 710-581-4766
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  618
;      TYPE:  AMINO ACID
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:
;      DESCRIPTION:  PROTEIN
;    HYPOTHETICAL:  YES
;    ORIGINAL SOURCE:
;      ORGANISM:  RAT
;      INDIVIDUAL ISOLATE:  SODIUM/IODIDE SYMPORTER
S-08-595-553A-2

  Query Match             7.3%;  Score 218;  DB 4;  Length 618;
  Best Local Similarity   20.8%;  Pred. No. 1.3e-12;
  Matches  125;  Conservative  110;  Mismatches  264;  Indels  102;  Gaps   23;

Qy      13 VFYLLILL---VGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYIN 69
Db      19 VFATMLLVSTGIGLWVGLARGGQRSADD----FFTGGRQLAAVPVGLSLAASFMSAVQVL 74

Qy      70 GT-AEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQIYGK--RM 126
Db      75 GVPAEAA---RYGLKFLWMCAGQLLNSLLTAFLFLPIFYRLGLTSTYQYLELRFSRAVRL 131

Qy     127 GGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGGLYSVAY 186
Db     132 CGTL--QYLVATMLYTGIVIYAPALILNQVTGLDIWASLLSTGIICTLYTTVGGMKAVVW 189

Qy     187 TDVVQLFCIFVGLWI----------SVPFALSHPAVADIGFTAVHAKYQKPWLGTV 232
Db     190 TDVFQVVVMLVGFWVILARGVILLGGPRNVLSLAQNHSRINLMDFDPDPRSRYTFWTFIV 249

Qy     233 DSSEVYSWLDSFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIG 292
Db     250 GGTLV--WLSMY------GVN-QAQVQRYVACHTEGKAKLALLVNQLGLFLIVASAACCG 300

Qy     293 AIGASTDWNQTAYGLPDPKTT---EEADMILPIVLQYL---CPVYISFFGLGAVSAAVMS 346
Db     301 IV------MFVYYKDCDPLLTGRISAPDQYMPLLVLDIFEDLPGVPGLF-LACAYSGTLS 353

Qy     347 SADSSILSASSMFARNIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLW 406
Db     354 TASTSINAMAAVTVEDLIKPRM-PGLAPRKLVFISKGLSFIYGSACLTVAALSSLLGGGV 412

Qy     407 YLSSDLVYIVIFPQLLCV----LFVKGTNTYGAVAGYVSGLFLRI---------TGGEPY 453
Db     413 LQGSFTVMGVISGPLLGAFTLGMLLPACNTPGVLSGLAAGLAVSLWVAVGATLYPPGEQT 472

Qy     454 LYLQP----------LIFYPGYYPDDNGIYNQKFP----------------FKTLAMV 485
Db     473 MGVLPTSAAGCTNDSVLLGPPGATNASNGIPSSGMDTGRPALADTFYAISYLYYGALGTL 532

Qy     486 TSFLTNICISYLAKYLFESGTLPPKLDVFDAVVARHSEENMDKTILVKNENIKLDELALV 545
Db     533 TTMLCGALISYLTGPTKRS-SLGPGLLWWD--LARQT-----ASVAPKEDTATLEE-SLV 583

Qy     546 K 546
Db     584 K 584

RESULT 6
US-09-134-001C-4510
; Sequence 4510, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOC
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4510
;   LENGTH: 530
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4510

  Query Match             6.9%;  Score 206.5;  DB 4;  Length 530;
  Best Local Similarity   22.7%;  Pred. No. 1.3e-11;
  Matches  117;  Conservative   89;  Mismatches  199;  Indels  111;  Gaps   26;

Qy       9 IAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWV----- 63
Db       4 ISIILFIMVVGGISFYAYLQSRKIKTSS--SDGYFMGGNSL----TGFTVASTIIMTNLS 57

Qy      64 --------GGGYING---------TAEAVYVPGYGLAWAQAP--IGYSLSLILGGLFFAK 104
Db      58 TEQIVGQNGQSYAQGMEVMAWEVTAAVAVVL----LAWVFLPKYLKYGVNTISEFLELRY 113

Qy     105 PMRSKGYVTMLDPFQQIYGKRMGGLLFIPALM-------GEMFWAAAIFSALGATISVII 157
Db     114 DTFTKRFVSILFIFTYV-------VSFLPVVLYSGSLVFNKMF---KVDEYLGVSSSTAV 163

Qy     158 DVDMHISVIISALIATLYTLVGGLYSVAYTDVVQLFCIFV-GLWISVPFALSHPAVADIG 216
Db     164 II---ISSII-GIIGIIYLFIGGLSLSAFSDSIYGMALIIGGLAITI---LGLGQLGDGN 216

Qy     217 FTAVHAKYQKP------WLGTVDSSEVYSWLDSFLLLMLGGI-PW---QAYFQRVLSSSS 266
Db     217 FLHGFDKIVQDTPEKLNGFGKVD-SDVVPWPTLFFGMFFNNLFFWCANQMIVQKALAAKN 275

Qy     267 ATYAQ----VLSFLAAFGCLVMAIPAIL-IGAIGASTDWNQTAYGLPDPKTTEEADMILP 321
```

```
Db    276 LKESQKGAIYLSLFKVFGPLITVLPGVVAFNYFNGSIDKSDNAY--------------P 320

Qy    322 IVLQYLCPVYISFFGL-GAV-SAAVMSSADSSILSASSMFARNIYQLSFRQNASDKEIVW 379

Db    321 ALVTSVLPEWA--FGLFGAVIFGAILSSFVGSLNSTTTLLTLDFYKPIFGKNKSDKHIAR 378

Qy    380 VMRITVFVFGASATAMA----LLTKTVYGLWYLSSDLVYIVIFPQLLCVLFVKGTNTYGA 435

Db    379 VGHIATVVIGVIVVALAPVISLFPSGLYAVVQQFNGVYSMPVLALILVAFFSKRTSKLGA 438

Qy    436 ---------VAGYVSGLFLRITGGEPYLYLQPLIFY 462

Db    439 KVTLFTHIILYAIISFVFTEIN----YLYTFSVLFF 470


RESULT 7
US-09-014-969-11
; Sequence 11, Application US/09014969
; Patent No. 5965397
; GENERAL INFORMATION:
;   APPLICANT: Jacobs, Kenneth
;   APPLICANT: McCoy, John M.
;   APPLICANT: LaVallie, Edward R.
;   APPLICANT: Racie, Lisa A.
;   APPLICANT: Merberg, David
;   APPLICANT: Treacy, Maurice
;   APPLICANT: Spaulding, Vikki
;   APPLICANT: Agostino, Michael J.
;   TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
;   TITLE OF INVENTION: ENCODING THEM
;   NUMBER OF SEQUENCES: 32
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Genetics Institute, Inc.
;     STREET: 87 CambridgePark Drive
;     CITY: Cambridge
;     STATE: MA
;     COUNTRY: U.S.A.
;     ZIP: 02140
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/014,969
;     FILING DATE:
;     CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Sprunger, Suzanne A.
;     REGISTRATION NUMBER: 41,323
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (617) 498-8284
;     TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 11:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 635 amino acids
;     TYPE: amino acid
;     STRANDEDNESS:
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-09-014-969-11

  Query Match             6.7%;  Score 200;  DB 2;  Length 635;
  Best Local Similarity   23.8%;  Pred. No. 7.5e-11;
  Matches  120;  Conservative   70;  Mismatches  191;  Indels  124;  Gaps   23;

Qy      3 FHVEGLIAIIVFYLLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATW 62

Db     24 FSIMDYVVFVLLLVLSLAIGLYHACR----GWGRHTVGELLMADRKMGCLPVALSLLATF 79

Qy     63 VGGGYINGTAEAVYVPG-----------YGL---AWAQAPIGYSLSLILG----GLFFAK 104

Db     80 QSAVAILRVPSEIYRFGTQYWFLRCCYFLGLLIPAHIFIPVFYRLHLTSAYEYLELRFNK 139

Qy    105 PMRSKGYVTMLDPFQQIVGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDVDMHIS 164

Db    140 TVRVCGTVTFI--FQMVI--YMGVVLYAPSL----------------ALNAVTGFDLWLS 179

Qy    165 VIISALIATLYTLVGGLYSVAYTDVVQLFCIFVG----------------LWISVPFAL 207

Db    180 VLALRIVCTVYTALGGLKAVIWTDVFQTLVMFLGQLAVIIVGSAKVGGLGRVW---AVAS 236

Qy    208 SHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSWLDSFLLLMLGGIPWQAYFQRVLSSSSA 267

Db    237 QHGRIS--GFELDPDPFVRHTFWTLAFGGV------FMMLSLYGVN-QAQVQRYLSSRTE 287

Qy    268 TYAQVLSFLAAF---------GCLVMAIPAILIGAIGASTDWNQTAYGLPDPKTTEEADM 318

Db    288 K-AAVLSCYAVFPFQQVSLCVGC--------LIGLV-------MFAYYQEYPMSIQQAQA 331

Qy    319 IL-PIVLQYLCPVYISFFGL-GAVSAAVMSSADSSILSASSMFARNIYQLSFR---QNAS 373

Db    332 APDQFVLYFVMDLLKGLPGLPGLFIACLFSGSLSTISSAFNSLATVTMEDLIRPWFPEFS 391

Qy    374 DKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDL-----VYIVIFPQ-------L 421

Db    392 EARAIMLSRGLAFGYGLLCLGMA----------YISSQMGPVLQAAISIFGMVGGPLLGL 441

Qy    422 LCV-LFVKGTNTYGAVAGYVSGLFL 445

Db    442 FCLGMFFPCANPPGAVVGLLAGLVM 466


RESULT 8
US-09-134-001C-4309
; Sequence 4309, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCC
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4309
;   LENGTH: 482
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4309

  Query Match             4.6%;  Score 138;  DB 4;  Length 482;
  Best Local Similarity   21.9%;  Pred. No. 5.7e-05;
  Matches  102;  Conservative   78;  Mismatches  176;  Indels  110;  Gaps   24;

Qy      1 MAFHVEGLIAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60

Db     49 MGGHAGGL-AIIIGWLITAIGMISLALVFQN--LTNERSD----------LDGGIYSYA 94

Qy     61 TWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQ 120

Db     95 QAGFGDFI---------GFASAW-----GYWFSAFLGNVAYATLLMSS-----IGNFFP 134

Qy    121 IYGKRMGGLLFIPALMGE-MFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVG 179

Db    135 IF---KGGNTFPSIIVASILLWSVHFLILKGVETAALINSIVTITKLIPILLVIICMIVA 191

Qy    180 -------------GLYSVAY----------TDVVQLFCIFVGLWISVPFALSHPAVAD 214

Db    192 FNFNTFRIGFFGMDGYGSLSFHFANTMSQVKSTMLVTVW-VFIGIEGAVVFSGRAKNKKD 250
```

```
Qy     215 IGFTAVHAKYQKPWLGTVDSSEVYSWLDSFLLLMLG-GIPWQAYFQRVLSSSSATYAQVL 273
           :|   |       :| :    :|     ||| :|  ||  | :  ::    : : ||:|
Db     251 VGTATV--------IGLISVLLIY-----FLLTVLAQGIVIQNHISKL---EAPSMAQIL 294

Qy     274 SFLAA-FGCLVMAIPAILIGAIGASTDWNQTAYGLP--------DPK-TTEEADMILPIV 323
           :::   :|   : |  ::|  :||   |   |  ||         ||   :|      |:
Db     295 AYIVGDWGATFVNI-GLIISVLGAWLGWTLLAGELPFIVAKDGLFPKWFAKENKNGAPMN 353

Qy     324 LQYLCPVYISFFGLGAVSAA-----VMSSADSSIL--SASSMFARNIYQLSFRQNASDKE 376
           ::  | :  | :  :           | | |:||     | | :  | : ::  |: |:
Db     354 ALFITNVLVQIFLISMLFTKSAYHFAFSLAASAILYPYMFSAFYQVKYTIEYKLTATPKQ 413

Qy     377 IVWVMRITVFVFG-----ASATAMALLTKTVYGLWYLSSDLVYIVI 417
           |:: |   ::      ||     |||     | |:   :||:|:
Db     414 --WIIGILASIYAIWLVYASGIDYLLLTM----LLYIPGIIVYVVV 453


RESULT 9
US-09-134-001C-3001
; Sequence 3001, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3001
;   LENGTH: 496
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3001

  Query Match             4.1%;  Score 123;  DB 4;  Length 496;
  Best Local Similarity   20.3%;  Pred. No. 0.0017;
  Matches  100;  Conservative   65;  Mismatches  177;  Indels  150;  Gaps   22;

Qy      49 IGLLVGGFTMTATWVGGGYINGTAEAVYVPGYGLA----WAQAPIG-YSLSLILGGLFFA 103
           |||:|      : :|||  |  ::     | |||     |    ||  ||: :   |
Db      35 IGLVIG------SMIGGGAFNIISDMGGQAG-GLAIIIGWIITAIGMISLAFVFQNLTNE 87

Qy     104 KPMRSKG-YVTMLDPFQQIYGKRMGGLLFIPALMGEMFWAAAIFSALG------------ 150
           :|    | |       |    |         : :| : :|  : ||:|
Db      88 RPDLDGGIYSYAQTGFGDFIGFSSAWGYWFAAFLGNVAYATLLMSAVGNFFPIFKGGNTL 147

Qy     151 ------------------------ATISVIIDV----DMHISVIISALIATLYTLVGGLY 182
                                    | |: |: |      : : :|   ::     |   | |
Db     148 PSIIIASILLWGVHFLILRGVETAAFINSIVTVAKLIPIFLVIICMIVVFNFSTFKSGFY 207

Qy     183 -----------------SVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQ 225
                            |  | :| :: :| |:  :| |:     :  |:|   |
Db     208 GMTSGSVGVFSWGDTMAQVKSTMLVTVW-VFTGIEGAVVFSGRAKSKKDVGTATV----- 261

Qy     226 KPWLGTVDSSEVYSWLDSFLLLMLG-GIPWQAYFQRVLSSSSATYAQVLSFLAAFGCLVM 284
           :| :    :|     ||: :|  |:  |    :: : |    ||||  :          |:
Db     262 ---IGLISVLVIY-----FLMTVLAQGVIHQNQISKLANPS---MAQVLEHIVGHWGSVL 310

Qy     285 AIPAILIGAIGASTDWNQTAYGLP----------------DPKTTEEADMILPIVLQYL 327
           ::|  :||   |   |  ||                    |    | :|  |::|
Db     311 VNIGLIISVLGAWLGWTLLAGELPFIVAKDGLFPKWFAKENKNKAPVNALIITNILVQLF 370

Qy     328 CPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFR-QNASDKEI-VWVMRITV 385
           ||     :   | | |:||   :: |   ||: :  || |   :  |:: |
Db     371 ---LISMLFTDSAYQFAFSLASSAILIPYTLSA--FYQVKYTIQNKSKANLKQWIIGIIA 425

Qy     386 FVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQLLCVLFVKGTNTYGAVAGYVSGLFL 445
                       ::| :|     |||      ||  : :           |: || :
Db     426 ---------------SIYTIW-----LVYAAGLDYLLLTMLL----------YIPGLLV 454

Qy     446 RITGGEPYLYLQ 457
                  | |:|
Db     455 -------YSYVQ 459


RESULT 10
US-09-134-001C-5370
; Sequence 5370, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCO
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5370
;   LENGTH: 499
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5370

  Query Match             4.0%;  Score 119;  DB 4;  Length 499;
  Best Local Similarity   20.8%;  Pred. No. 0.0043;
  Matches  100;  Conservative   68;  Mismatches  180;  Indels  132;  Gaps   21;

Qy      19 LLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYINGTAEAVYVP 78
           :| | |     | ||               ||:::|   |    |  | :    |  |
Db      40 ILPGDW----IKQSGPI----------ASSIGIVIGALLMILIAVSYGAL---VEKFPVS 82

Qy      79 GYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQIYGKRMGGLLFIPALMGE 138
           |   |::    |  :|      ||:    : ||| ::      :        | |  |:
Db      83 GGAFAFSFLSFGRYVS------FFSSWFLTFGYVCVVALNATAFS------LLIKFLLPN 130

Qy     139 MFWAAAIFSALGATISVIIDVDMHISVIISALIATLYTLVGG----LYSVAYTDVVQLFC 194
           :     :::  |  : :   |   : :|:  ||      | |      : ||   || |
Db     131 VLNNGKLYTVAGWDVYITEIVIATVLLIVFMLITIRGASVSGSLQYYFCVAMVLVVAL-- 188

Qy     195 IFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSWLDSFLLLMLGGIPW 254
           :|:| : |  |:|||                :| | :||      |  | :::::   ||
Db     189 MFIGSFFSSHFSLSH--------------LEP-LASVDK----GWFQS-IIMIVSIAPW 227

Qy     255 QAY--FQRV--------------------LSSSSATYAQVLSFLA-------------- 277
            ||  |  :                    | ::| ||  :| :
Db     228 -AYVGFDNIPQTAEEFNFSPNKTFKLIVYSLLAASLTYVVMLLYTGWLSTQATSLNGNLW 286

Qy     278 --------AFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKTTEEADMILPIVLQYL-- 327
                   ||| : :|: |: |  :|  |  |           :     |:| |   |
Db     287 LTGAVTQDAFGFIGLAVLAVAI-IMGIFTGLNGFLMSSSRLLFSMGRSGIMPTVFSKLHS 345

Qy     328 ---CPVYISFFGLGAVSAAV-------------MSSADSSI------LSASSMFA----R 361
              | |::   | |||                 |||   ||      |||: :|:
Db     346 KHKTP-YVAIIFLVAVSLIAPWLGRTALTWIVDMSSTGVSIAYFITCLSATKLFSFNKQS 404

Qy     362 NIYQLSFRQNASDKEIVWVMRITVFVFGASATAMALLTKTVYGLWYLSSDLVYIVIFPQL 421
           | |   ::        ||  : : : :   |  |::: :    |:| :   : :|:  |:|
Db     405 NTYAPVYKIFGIIGSIVSFVFLCLLLIPGSPAALSIPSYIALGIWLVIGLIFFIIRLPKL 464


RESULT 11
US-08-677-734A-12
; Sequence 12, Application US/08677734A
; Patent No. 5871919
;  GENERAL INFORMATION:
;    APPLICANT:  Brant, Steven R.
;    APPLICANT:  Yun, Chris C.H.
```

```
;    APPLICANT:  Donowitz, Mark
;    APPLICANT:  Tse, Chung-Ming
;    TITLE OF INVENTION:  Cloning, Tissue Distribution, and
;    TITLE OF INVENTION:  Functional Analysis Of The Human Na+/H+ Exchanger Isoform,
;    TITLE OF INVENTION:  NHE3.
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/677,734A
;      FILING DATE:  10-JUL-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fordis, Jean B.
;      REGISTRATION NUMBER:  32,984
;      REFERENCE/DOCKET NUMBER:  05387.0043-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  832 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-677-734A-12

  Query Match             3.9%;  Score 116;  DB 2;  Length 832;
  Best Local Similarity   19.4%;  Pred. No. 0.019;
  Matches   86;  Conservative   58;  Mismatches  134;  Indels  166;  Gaps   19;

Qy       9 IAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYI 68
Db      66 LAKIVFHL------------SHKVTSVVPESALLIV----LGLVLGGIVLAADHIASFTL 109

Qy      69 NGTAEAVYV-------PGY--------------------GLAWAQAPIGYSLSLILGGLF 101
Db     110 TPTVFFFYLLPPIVLDAGYFMPNRLFFSNLGSILLYAVVGTVWNAATTGLSL-------- 161

Qy     102 FAKPMRSKGYVTMLDPFQQIYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDV-- 159
Db     162 --------------------YG------VFLSGIMGELKIGLLDFLLFGSLIAAVDPVAV 195

Qy     160 -----DMHI-----------SVIISALIATLYTLVGGLYSVAYTDVVQLFCI------FV 197
Db     196 LAVFEEVHVNEVLFIIVFGESLLNDAVTVVLYNVFQSFVTLGGDKVTGVDCVKGIVSFFV 255

Qy     198 ----GLWISVPFAL---------SHPAVADIGFTAV--HAKYQKPWLGTVDSSEVYSWLD 242
Db     256 VSLGGTLVGVVFAFLLSLVTRFTKHVRVIEPGFVFIISYLSYL--------TSEMLS-LS 306

Qy     243 SFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGC-----LVMAIPAI-------- 289
Db     307 SILAITFCGICCQKYVKANISEQSATTVRYTMKMLASGAETIIFMFLGISAVDPLIWTWN 366

Qy     290 ------------LIGAIGASTD-WNQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFG 336
Db     367 TAFVRLTLLFVSVFRAIGVVLQTWLLNRYRM-------------VQLELIDQVVMSYGG 412

Qy     337 L-GAVSAAVMSSADSSILSASSMF 359
```

```
Db     413 LRGAVAFALVALLDGNKVKEKNLF 436


RESULT 12
US-09-097-053-12
; Sequence 12, Application US/09097053
; Patent No. 6392025
;  GENERAL INFORMATION:
;    APPLICANT:  Brant, Steven R.
;    APPLICANT:  Yun, Chris C.H.
;    APPLICANT:  Donowitz, Mark
;    APPLICANT:  Tse, Chung-Ming
;    TITLE OF INVENTION:  Cloning, Tissue Distribution, and
;    TITLE OF INVENTION:  Functional Analysis Of The Human Na+/H+ Exchanger Isoform,
;    TITLE OF INVENTION:  NHE3.
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/097,053
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/677,734
;      FILING DATE:  10-JUL-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fordis, Jean B.
;      REGISTRATION NUMBER:  32,984
;      REFERENCE/DOCKET NUMBER:  05387.0043-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  832 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-097-053-12

  Query Match             3.9%;  Score 116;  DB 4;  Length 832;
  Best Local Similarity   19.4%;  Pred. No. 0.019;
  Matches   86;  Conservative   58;  Mismatches  134;  Indels  166;  Gaps   19;

Qy       9 IAIIVFYLLILLVGIWAAWRTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTATWVGGGYI 68
Db      66 LAKIVFHL------------SHKVTSVVPESALLIV----LGLVLGGIVLAADHIASFTL 109

Qy      69 NGTAEAVYV-------PGY--------------------GLAWAQAPIGYSLSLILGGLF 101
Db     110 TPTVFFFYLLPPIVLDAGYFMPNRLFFSNLGSILLYAVVGTVWNAATTGLSL-------- 161

Qy     102 FAKPMRSKGYVTMLDPFQQIYGKRMGGLLFIPALMGEMFWAAAIFSALGATISVIIDV-- 159
Db     162 --------------------YG------VFLSGIMGELKIGLLDFLLFGSLIAAVDPVAV 195

Qy     160 -----DMHI-----------SVIISALIATLYTLVGGLYSVAYTDVVQLFCI------FV 197
Db     196 LAVFEEVHVNEVLFIIVFGESLLNDAVTVVLYNVFQSFVTLGGDKVTGVDCVKGIVSFFV 255

Qy     198 ----GLWISVPFAL---------SHPAVADIGFTAV--HAKYQKPWLGTVDSSEVYSWLD 242
```

```
Db    256 VSLGGTLVGVVFAFLLSLVTRFTKHVRVIEPGFVFIISYLSYL--------TSEMLS-LS 306

Qy    243 SFLLLMLGGIPWQAYFQRVLSSSSATYAQVLSFLAAFGC-----LVMAIPAI-------- 289

Db    307 SILAITFCGICCQKYVKANISEQSATTVRYTMKMLASGAETIIFMFLGISAVDPLIWTWN 366

Qy    290 -----------LIGAIGASTD-WNQTAYGLPDPKTTEEADMILPIVLQYLCPVYISFFG 336

Db    367 TAFVRLTLLFVSVFRAIGVVLQTWLLNRYRM-------------VQLELIDQVVMSYGG 412

Qy    337 L-GAVSAAVMSSADSSILSASSMF 359

Db    413 LRGAVAFALVALLDGNKVKEKNLF 436


RESULT 13
US-09-068-195-24
; Sequence 24, Application US/09068195B
; Patent No. 6140078
; GENERAL INFORMATION:
;  APPLICANT: Sanders, Jan W.
;  APPLICANT: Ledeboer, Adrianus M.
;  APPLICANT: Venema, Gerard
;  APPLICANT: Kok, Jan
;  TITLE OF INVENTION: Salt-Inducible Promoter Derivable from a Lactic Acid
;  TITLE OF INVENTION: Bacterium, and Its Use in a Lacctic Acid Bacterium for
;  TITLE OF INVENTION: Production of Desired Protein
;  FILE REFERENCE: Sanders-60113/0252227
;  CURRENT APPLICATION NUMBER: US/09/068,195B
;  CURRENT FILING DATE: 1998-07-29
;  EARLIER APPLICATION NUMBER: PCT/EP97/04755
;  EARLIER FILING DATE: 1997-08-20
;  EARLIER APPLICATION NUMBER: EP 97200744/7
;  EARLIER FILING DATE: 1997-03-13
;  EARLIER APPLICATION NUMBER: EP 96202444/4
;  EARLIER FILING DATE: 1996-09-05
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
;  LENGTH: 503
;  TYPE: PRT
;  ORGANISM: L. lactis MG1363
US-09-068-195-24

  Query Match             3.8%;  Score 113.5;  DB 4;  Length 503;
  Best Local Similarity   20.9%;  Pred. No. 0.015;
  Matches  119;  Conservative   83;  Mismatches  207;  Indels  161;  Gaps   28;

Qy     52 LVGGFTMTATWVGGGYINGTAEAVYVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGY 111

Db      8 LFGFFALTASMVLTVYEYPT-----------FATSKLHLVFFLLLGGLLWFLPV----- 50

Qy    112 VTMLDPFQQIYGKRMGGLL-FIPALMGEMFWAAAIFSALGATISVIIDVDMHISVIISAL 170

Db     51 ALCAAEMATVEGWKNGGIFSWVSQTLGERFGFAAIF-------------FQWFQITVGF 96

Qy    171 IATLYTLVGGL-----YSVAYTD----VVQLFCIFVGLWISVPFALSHPAVADIGFTAVH 221

Db     97 VTMIYFILGALSYVLNFQALNTDPLIKFIGLLIIFWGLTFS----------QLGGTQRT 145

Qy    222 AKYQKPW--LGTVDSSEVYSWLDSFLLLMLGG----IPWQAYFQRVLSSSSATYAQVLSF 275

Db    146 AKLVKAGFVVGIVIPSVILFGLAA--AYFIGGNPIEIPINSHAFVPDFSQVSTLVVFVSF 203

Qy    276 LAAF---------------------GCLVMAIPAILIGAIGASTDWNQTAYGLPDPKTT 313

Db    204 ILAYMGVEASASHINELENPKRNYPLAMILLVILAISLDAIGGFS----VAAVIPQKELS 259

Qy    314 EEADMI---LPIVLQY---------LCPVYISFFGLGAVSAAVMSSADSSILSASSMFA- 360

Db    260 LSAGVIQTFQTLILHFNHHLGWLVKVIALMIAFGVMGEVSSWVVG-------PSRGMFAA 312

Qy    361 --RNIYQLSFRQNASDK---EIVWVMRITVFVFGASAT----------AMALLTKTVYG 404

Db    313 AQRGLLPKFLRKTNTHEVPVPLVMIQGIIVTLWGAVLTFGGGGNNLSFLVAISLTVVIYL 372

Qy    405 LWYLSSDLVYIVIFPQLLCVLFVKGTN---TYGAVAGYVSGLFLRITGGEPYLYLQPLIF 461

Db    373 VGYL---LFFIVYF-----VLIYKKQNLKRTYN-VPGKIIG--KTIIAGIGFL-LSIFAL 420

Qy    462 YPGYYPDDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYL--------------FESGTL 507

Db    421 FISFVPPASIAKNETHTYQMILLI-SFVVTAILPFIIYELHDKKGHDTIEEPTHFKAGDV 479

Qy    508 PPKLDVFDAVVARHSEENMDKTILVKNENI 537

Db    480 NPA--IYPAARGEHH-------IIKKEEHI 500


RESULT 14
US-08-295-814E-11
; Sequence 11, Application US/08295814E
; Patent No. 5658786
;  GENERAL INFORMATION:
;    APPLICANT:  Smith, Kelli E.
;    APPLICANT:  Borden, Laurence A.
;    APPLICANT:  Hartig, Paul R.
;    APPLICANT:  Weinshank, Richard L.
;    TITLE OF INVENTION:  DNA ENCODING TAURINE AND GABA
;    TITLE OF INVENTION:  TRANSPORTERS AND USES THEREOF
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/295,814E
;      FILING DATE:  DECEMBER 19, 1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40558-B-PCT-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-278-0400
;      TELEFAX:  212-391-0525
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  599 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-295-814E-11

  Query Match             3.7%;  Score 110;  DB 1;  Length 599;
  Best Local Similarity   17.8%;  Pred. No. 0.044;
  Matches   92;  Conservative   59;  Mismatches  146;  Indels  220;  Gaps   24;

Qy     29 TKNSGSAE-----ERSEAIIVGGRDIGLLV------GGFTMTATWVGGGYINGTAEAVYV 77

Db      3 TDNSKVADGQISTEVSEAPVASDKPKTLVVKVQKKAGDLPDRDTWKG------------ 49

Qy     78 PGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQIYGKRMGGLLFIPALMG 137

Db     50 ---RFDFLMSCVGYAIG--LGNVW--------------RFPYLCGKNGGGAFLIPYFLT 89
```

```
Qy    138 EMFWAAAIF---SALGATISVIIDVDMHISVIISALIATLYTLVGGLYSVAYTDVVQLFC 194
          :|    :|    :||                          || :|||
Db     90 LIFAGVPLFLLECSLGQ----------------------YTSIGGL------------ 113

Qy    195 IFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSWLDSFLLLMLGGIPW 254
             |:|   |          :|  |    :                ||: : :::   | |
Db    114 ---GVWKLAPM------FKGVGLAAAVLSF--------------WLNIYYIVI---ISW 146

Qy    255 QAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKTTE 314
            |:                          |  :   :  | |      :|  |:
Db    147 AIYY----------------------------LYNSFTTTLPWKQ----CDNPWNTD 171

Qy    315 EADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASD 374
                              |     |:   |:||          : ||::|::   :  |
Db    172 RC----------------FSNYSLVNTTNMTSA------VVEFWERNMHQMT---DGLD 205

Qy    375 K--EIVWVMRITVFVFGASATAMALLTKTVYGLW----------YLSSDLVYIVIFPQLL 422
          |  :| | : ||:      | |  |:    : :|          | |:   ||     :|
Db    206 KPGQIRWPLAITL------AIAWVLV---YFCIWKGVGWTGKVVYFSATYPYI-----ML 251

Qy    423 CVLFVKGTNTYGAVAG---YVSGLFLRITGGEPYLYLQPLIFYP------------GYYP 467
           :|| :|    ||  |   |::  | :::  | :|     ||:                |
Db    252 IILFFRGVTLPGAKEGILFYITPNFRKLSDSEVWLDAATQIFFSYGLGLGSLIALGSYNS 311

Qy    468 DDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFES 504
          | :|           :     | | |  | ::  |
Db    312 FHNNVYRDS--------IIVCCINSCTSMFAGFVIFS 340
```

```
RESULT 15
US-08-240-783B-4
; Sequence 4, Application US/08240783B
; Patent No. 5756348
;  GENERAL INFORMATION:
;    APPLICANT:  Smith, Kelli
;    APPLICANT:  Borden, Laurence A.
;    APPLICANT:  Branchek, Theresa
;    APPLICANT:  Hartig, Paul R.
;    APPLICANT:  Weinshank, Richard L.
;    TITLE OF INVENTION:  DNA ENCODING A GLYCINE TRANSPORTER AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/240,783B
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  1795/39875-A-PCT-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212)391-0525
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  599 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  N
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  RAT GABA TRANSPORTER (GAT-1)
US-08-240-783B-4

  Query Match             3.7%;  Score 110;  DB 1;  Length 599;
  Best Local Similarity   17.8%;  Pred. No. 0.044;
  Matches   92;  Conservative   59;  Mismatches  146;  Indels  220;  Gaps   24;

Qy     29 TKNSGSAE-----ERSEAIIVGGRDIGLLV------GGFTMTATWVGGGYINGTAEAVYV 77
          | ||  |:     | ||| :   :   |:|      |       || |
Db      3 TDNSKVADGQISTEVSEAPVASDKPKTLVVKVQKKAGDLPDRDTWKG------------ 49

Qy     78 PGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGYVTMLDPFQQIYGKRMGGLLFIPALMG 137
                :  : :||::   || ::                |  : ||  ||   ||  :
Db     50 ---RFDFLMSCVGYAIG--LGNVW--------------RFPYLCGKNGGGAFLIPYFLT 89

Qy    138 EMFWAAAIF---SALGATISVIIDVDMHISVIISALIATLYTLVGGLYSVAYTDVVQLFC 194
          :|    :|    :||                          || :|||
Db     90 LIFAGVPLFLLECSLGQ----------------------YTSIGGL------------ 113

Qy    195 IFVGLWISVPFALSHPAVADIGFTAVHAKYQKPWLGTVDSSEVYSWLDSFLLLMLGGIPW 254
             |:|   |          :|  |    :                ||: : :::   | |
Db    114 ---GVWKLAPM------FKGVGLAAAVLSF--------------WLNIYYIVI---ISW 146

Qy    255 QAYFQRVLSSSSATYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTAYGLPDPKTTE 314
            |:                          |  :   :  | |      :|  |:
Db    147 AIYY----------------------------LYNSFTTTLPWKQ----CDNPWNTD 171

Qy    315 EADMILPIVLQYLCPVYISFFGLGAVSAAVMSSADSSILSASSMFARNIYQLSFRQNASD 374
                              |     |:   |:||          : ||::|::   :  |
Db    172 RC----------------FSNYSLVNTTNMTSA------VVEFWERNMHQMT---DGLD 205

Qy    375 K--EIVWVMRITVFVFGASATAMALLTKTVYGLW----------YLSSDLVYIVIFPQLL 422
          |  :| | : ||:      | |  |:    : :|          | |:   ||     :|
Db    206 KPGQIRWPLAITL------AIAWVLV---YFCIWKGVGWTGKVVYFSATYPYI-----ML 251

Qy    423 CVLFVKGTNTYGAVAG---YVSGLFLRITGGEPYLYLQPLIFYP------------GYYP 467
           :|| :|    ||  |   |::  | :::  | :|     ||:                |
Db    252 IILFFRGVTLPGAKEGILFYITPNFRKLSDSEVWLDAATQIFFSYGLGLGSLIALGSYNS 311

Qy    468 DDNGIYNQKFPFKTLAMVTSFLTNICISYLAKYLFES 504
          | :|           :     | | |  | ::  |
Db    312 FHNNVYRDS--------IIVCCINSCTSMFAGFVIFS 340

Search completed: March 31, 2003, 18:38:34
Job time : 43 secs
```